

Field of the Invention

~~1~~(C) This invention relates to proteins having serine/threonine kinase domains, corresponding nucleic acid molecules, and their use.

Background of the Invention

The transforming growth factor- β (TGF- β) superfamily consists of a family of structurally-related proteins, including three different mammalian isoforms of TGF- β (TGF- β 1, β 2 and β 3), activins, inhibins, müllerian-inhibiting substance and bone morphogenic proteins (BMPs) (for reviews see Roberts and Sporn, (1990) Peptide Growth Factors and Their Receptors, Pt.1, Sporn and Roberts, eds. (Berlin: Springer - Verlag) pp 419-472; Moses et al (1990) Cell 63, 245-247). The proteins of the TGF- β superfamily have a wide variety of biological activities. TGF- β acts as a growth inhibitor for many cell types and appears to play a central role in the regulation of embryonic development, tissue regeneration, immuno-regulation, as well as in fibrosis and carcinogenesis (Roberts and Sporn (199) see above).

Activins and inhibins were originally identified as factors which regulate secretion of follicle-stimulating hormone secretion (Vale et al (1990) Peptide Growth Factors and Their Receptors, Pt.2, Sporn and Roberts, eds. (Berlin: Springer-Verlag) pp.211-248). Activins were also shown to induce the differentiation of haematopoietic progenitor cells (Murata et al (1988) Proc. Natl. Acad. Sci. USA 85, 2434 - 2438; Eto et al (1987) Biochem. Biophys. Res. Commun. 142, 1095-1103) and induce mesoderm formation in *Xenopus* embryos (Smith et al (1990) Nature 345, 729-731; van den Eijnden-Van Raaij et al (1990) Nature 345, 732-734).

35 BMPs or osteogenic proteins which induce the formation of bone and cartilage when implanted subcutaneously (Wozney et al (1988) Science 242, 1528-1534), facilitate neuronal

[illegible]

5
10

15
20
25
30

35

Among the type II receptors for the TGF- β superfamily of proteins, the cDNA for the activin type II receptor (ActRII) was the first to be cloned (Mathews and Vale (1991) Cell 65, 973-982). The predicted structure of the receptor was shown to be a transmembrane protein with an intracellular serine/threonine kinase domain. The activin receptor is related to the C. elegans daf-1 gene product, but the ligand is currently unknown (Georgi et al (1990) Cell 61, 635-645). Thereafter, another form of the activin type II receptor (activin type IIB receptor), of which there are different splicing variants (Mathews et al (1992), Science 225, 1702-1705; Attisano et al (1992) Cell 68, 97-108), and the TGF- β type II receptor (TBRII) (Lin et al (1992) Cell 68, 775-785) were cloned, both of which have putative serine/threonine kinase domains.

The present invention involves the discovery of related novel peptides, including peptides having the activity of those defined herein as SEQ ID Nos. 2, 4, 8, 10, 12, 14, 16 and 18. Their discovery is based on the realisation that receptor serine/threonine kinases form a new receptor family, which may include the type II receptors for other proteins in the TGF- β superfamily. To ascertain whether there were other members of this family of receptors, a protocol was designed to clone ActRII/daf I related cDNAs. This approach made use of the polymerase chain reaction (PCR), using degenerate primers based upon the amino-acid sequence similarity between kinase domains of the mouse activin type II receptor and daf-I gene products.

This strategy resulted in the isolation of a new family of receptor kinases called Activin receptor like kinases (ALK's) 1-6. These cDNAs showed an overall 33-39% sequence similarity with ActRII and TGF- β type II receptor and 40-92% sequence similarity towards each other in the kinase domains.

Soluble receptors according to the invention comprise at least predominantly the extracellular domain. These can be selected from the information provided herein, prepared in conventional manner, and used in any manner associated with the invention.

Antibodies to the peptides described herein may be raised in conventional manner. By selecting unique sequences of the peptides, antibodies having desired specificity can be obtained.

The antibodies may be monoclonal, prepared in known manner. In particular, monoclonal antibodies to the extracellular domain are of potential value in therapy.

Products of the invention are useful in diagnostic methods, e.g. to determine the presence in a sample for an analyte binding therewith, such as in an antagonist assay. Conventional techniques, e.g. an enzyme-linked immunosorbent assay, may be used.

Products of the invention having a specific receptor activity can be used in therapy, e.g. to modulate conditions associated with activin or TGF- β activity. Such conditions include fibrosis, e.g. liver cirrhosis and pulmonary fibrosis, cancer, rheumatoid arthritis and glomeronephritis.

Brief Description of the Drawings

Figure 1 shows the alignment of the serine/threonine (S/T) kinase domains (I-VIII) of related receptors from transmembrane proteins, including embodiments of the present invention. The nomenclature of the subdomains is accordingly to Hanks *et al* (1988).

Figures 2A to 2D shows the sequences and characteristics of the respective primers used in the

SECRETED/275060

cln22

Figure 3 is a comparison of the amino-acid sequences of human activin type II receptor (Act R-II), mouse activin type IIB receptor (Act R-IIB), human TGF- β type II receptor (TBR-II), human TGF- β type I receptor (ALK-5), human activin receptor type IA (ALK-2), and type IB (ALK-4), ALKS 1 & 3, and mouse ALK-6.

Figure 4 shows, schematically, the structures for Daf-1, Act R-II, Act R-IIB, TBR-II, TBR-I/ALK-5, ALK's -1, -2 (Act RIA), -3, -4 (Act RIB) & -6.

Figure 5 shows the sequence alignment of the cysteine-rich domains of the ALKs, TBR-II, Act R-II, Act R-IIB and *daf-1* receptors. (position 30-110 of SEQ ID NO:34) (position 29-106 of SEQ ID NO:34) (position 47-143 of SEQ ID NO:34)

Figure 6 is a comparison of kinase domains of serine/threonine kinases, showing the percentage amino-acid identity of the kinase domains.

Figure 7 shows the pairwise alignment relationship between the kinase domains of the receptor serine/threonine kinases. The dendrogram was generated using the Jotun-Hein alignment program (Hein (1990) Meth. Enzymol. 181, 626-645).

Brief Description of the Sequence Listings

Sequences 1 and 2 are the nucleotide and deduced amino-acid sequences of cDNA for hALK-1 (clone HP57).

Sequences 3 and 4 are the nucleotide and deduced amino-acid sequences of cDNA for hALK-2 (clone HP53).

Sequences 5 and 6 are the nucleotide and deduced amino-acid sequences of cDNA for hALK-3 (clone ONF5).

Sequences 7 and 8 the nucleotide and deduced amino-acid sequences of cDNA for hALK-4 (clone 11H8), complemented with PCR product encoding extracellular domain.

Sequences 9 and 10 are the nucleotide and deduced amino-acid sequences of cDNA for hALK-5 (clone EMBL).

Sequences 11 and 12 are the nucleotide and deduced amino-acid sequences of cDNA for mALK-1 (clone AM6).

Sequences 15 and 16 are the nucleotide and deduced
5 amino-acid sequences of cDNA for MALK-4 (clone 8a1).

Sequences 17 and 18 are the nucleotide and deduced amino-acid sequences of cDNA for MALK-6 (clone ME-6).

Sequence 19 (B1-S) is a sense primer, extracellular domain, cysteine-rich region, BamHI site at 5' end, 28-mer, 64-fold degeneracy.

Sequence 20 (B3-S) is a sense primer, kinase domain II, BamHI site at 5' end, 25-mer, 162-fold degeneracy.

Sequence 21 (B7-S) is a sense primer, kinase domain
VIB, S/T kinase specific residues, BamHI site at 5' end,
15 24-mer, 288-fold degeneracy.

Sequence 22 (E8-AS) is an anti-sense primer, kinase domain, S/T kinase-specific residues EcoRI site at 5' end, 20-mer, 18-fold degeneracy.

Sequence 23 is an oligonucleotide probe.

20 Sequence 24 is a 5' primer.

Sequence 25 is a 3' primer.

Sequence 26 is a consensus sequence in Subdomain I.

Sequences 27 and 28 are novel sequence motifs in Subdomain VIB.

25 Sequence 29 is a novel sequence motif in Subdomain
VIII.

Description of the Invention

As described in more detail below, nucleic acid sequences have been isolated, coding for a new sub-family of serine/threonine receptor kinases. The term nucleic acid molecules as used herein refers to any sequence which codes for the murine, human or mammalian form, amino-acid sequences of which are presented herein. It is understood that the well known phenomenon of codon degeneracy provides for a great deal of sequence variation and all such varieties are included within the scope of this invention.

5

10

25

30

35

5

10

15

20

25

30

35

(Promega, Madison, Wisconsin, U.S.A.) as described by the manufacturers, or purified through an oligo (dT)-cellulose column as described by Aviv and Leder (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412. The isolated mRNA was used for the synthesis of random primed (Amersham) cDNA, that was used to make a λ gt10 library with 1×10^5 independent cDNA clones using the Riboclone cDNA synthesis system (Promega) and λ gt10 in vitro packaging kit (Amersham) according to the manufacturers' procedures. An amplified oligo (dT) primed human placenta λ ZAPII cDNA library of 5×10^5 independent clones was used. Poly (A)⁺ RNA isolated from AG1518 human foreskin fibroblasts was used to prepare a primary random primed λ ZAPII cDNA library of 1.5×10^6 independent clones using the RiboClone cDNA synthesis system and Gigapack Gold II packaging extract (Stratagene). In addition, a primary oligo (dT) primed human foreskin fibroblast λ gt10 cDNA library (Claesson-Welsh et al (1989) Proc. Natl. Acad. Sci. USA. 86 4917-4912) was prepared. An amplified oligo (dT) primed HEL cell λ gt11 cDNA library of 1.5×10^6 independent clones (Poncz et al (1987) Blood 69 219-223) was used. A twelve-day mouse embryo λ EX10x cDNA library was obtained from Novagen (Madison, Wisconsin, U.S.A.); a mouse placenta λ ZAPII cDNA library was also used.

25 Generation of cDNA Probes by PCR

For the generation of cDNA probes by PCR (Lee et al (1988) Science 239, 1288-1291) degenerate PCR primers were constructed based upon the amino-acid sequence similarity between the mouse activin type II receptor (Mathews and Vale (1991) Cell 65, 973-982) and daf-1 (George et al (1990) Cell 61, 635-645) in the kinase domains II and VIII. Figure 1 shows the aligned serine/threonine kinase domains (I-VIII), of four related receptors of the TGF- β superfamily, i.e. hTBR-II, mActR-IIB, mActR-II and the daf-1 gene product, using the nomenclature of the subdomains according to Hanks et al (1988) Science 241, 45-52.

In order to facilitate directional subcloning, restriction enzyme sites were included at the 5' end of the primers, with a GC clamp, which permits efficient restriction enzyme digestion. The primers utilised are shown in Figure 2. Oligonucleotides were synthesized using Gene assembler plus (Pharmacia - LKB) according to the manufacturers instructions.

25 The mRNA prepared from HEL cells as described above
was reverse-transcribed into cDNA in the presence of 50 mM
Tris-HCl, pH 8.3, 8 mM MgCl₂, 30 mM KCl, 10 mM
dithiothreitol, 2mM nucleotide triphosphates, excess oligo
(dT) primers and 34 units of AMV reverse transcriptase at
30 42°C for 2 hours in 40 µl of reaction volume.
Amplification by PCR was carried out with a 7.5% aliquot (3
µl) of the reverse-transcribed mRNA, in the presence of 10
mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 M MgCl₂, 0.01% gelatin,
0.2 mM nucleotide triphosphates, 1 µM of both sense and
35 antisense primers and 2.5 units of Taq polymerase (Perkin
Elmer Cetus) in 100 µl reaction volume. Amplifications
were performed on a thermal cycler (Perkin Elmer Cetus)

using the following program: first 5 thermal cycles with denaturation for 1 minute at 94°C, annealing for 1 minute at 50°C, a 2 minute ramp to 55°C and elongation for 1 minute at 72°C, followed by 20 cycles of 1 minute at 94°C, 30 seconds at 55°C and 1 minute at 72°C. A second round of PCR was performed with 3 µl of the first reaction as a template. This involved 25 thermal cycles, each composed of 94°C (1 min), 55°C (0.5 min), 72°C (1 min).

General procedures such as purification of nucleic acids, restriction enzyme digestion, gel electrophoresis, transfer of nucleic acid to solid supports and subcloning were performed essentially according to established procedures as described by Sambrook *et al.*, (1989), Molecular cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Laboratory (Cold Spring Harbor, New York, USA).

Samples of the PCR products were digested with BamHI and EcoRI and subsequently fractionated by low melting point agarose gel electrophoresis. Bands corresponding to the approximate expected sizes, (see Table 1: ≈460 bp for primer pair B3-S and E8-AS and ≈ 140 bp for primer pair B7-S and E8-AS) were excised from the gel and the DNA was purified. Subsequently, these fragments were ligated into pUC19 (Yanisch-Perron *et al.* (1985) Gene 33, 103-119), which had been previously linearised with BamHI and EcoRI and transformed into *E. coli* strain DH5α using standard protocols (Sambrook *et al.*, *supra*). Individual clones were sequenced using standard double-stranded sequencing techniques and the dideoxynucleotide chain termination method as described by Sanger *et al.* (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467, and T7 DNA polymerase.

Employing Reverse Transcriptase PCR on HEL mRNA with the primer pair B3-S and E8-AS, three PCR products were obtained, termed 11.1, 11.2 and 11.3, that corresponded to novel genes. Using the primer pair B7-S and E8-AS, an additional novel PCR product was obtained termed 5.2.

TABLE 1

NAME OF PCR PRODUCT	PRIMERS	INSERT SIZE (bp)	SIZE OF DNA FRAGMENT IN hActRII/hTBR-II CLONES (bp)	SEQUENCE IDENTITY WITH SEQUENCE hActRII/hTBR-II (%)	SEQUENCE IDENTITY BETWEEN hActRII and TBR-II (%)
11.1	B3-S/E8-AS	460	460	46/40	42
11.2	B3-S/E8-AS	460	460	49/44	47
11.3	B3-S/E8-AS	460	460	44/36	48
11.29	B3-S/E8-AS	460	460	ND/100	ND
9.2	B1-S/E8-AS	800	795	100/ND	ND
5.2	B7-S/E8-AS	140	143	40/38	60

15 Isolation of cDNA Clones

The PCR products obtained were used to screen various cDNA libraries described supra. Labelling of the inserts of PCR products was performed using random priming method (Feinberg and Vogelstein (1983) Anal. Biochem, 132 6-13) using the Megaprime DNA labelling system (Amersham). The oligonucleotide derived from the sequence of the PCR product 5.2 was labelled by phosphorylation with T4 polynucleotide kinase following standard protocols (Sambrook et al, supra). Hybridization and purification of positive bacteriophages were performed using standard molecular biological techniques.

The double-stranded DNA clones were all sequenced using the dideoxynucleotide chain-termination method as described by Sanger et al, supra, using T7 DNA polymerase (Pharmacia - LKB) or Sequenase (U.S. Biochemical Corporation, Cleveland, Ohio, U.S.A.). Compressions of nucleotides were resolved using 7-deaza-GTP (U.S. Biochemical Corp.) DNA sequences were analyzed using the DNA STAR computer program (DNA STAR Ltd. U.K.). Analyses of the sequences obtained revealed the existence of six

distinct putative receptor serine/threonine kinases which have been named ALK 1-6.

To clone cDNA for ALK-1 the oligo (dT) primed human placenta cDNA library was screened with a radiolabelled insert derived from the PCR product 11.3; based upon their restriction enzyme digestion patterns, three different types of clones with approximate insert sizes. of 1.7 kb, 2 kb & 3.5 kb were identified. The 2 kb clone, named HP57, was chosen as representative of this class and subjected to complete sequencing. Sequence analysis of ALK-1 revealed a sequence of 1984 nucleotides including a poly-A tail (SEQ ID No. 1). The longest open reading frame encodes a protein of 503 amino-acids^{SEQ ID No. 2} with high sequence similarity to receptor serine/threonine kinases (see below). The first methionine codon, the putative translation start site, is at nucleotide 283-285 and is preceded by an in-frame stop codon. This first ATG is in a more favourable context for translation initiation (Kozak (1987) Nucl. Acids Res., 15, 8125-8148) than the second and third in-frame ATG at nucleotides 316-318 and 325-327. The putative initiation codon is preceded by a 5' untranslated sequence of 282 nucleotides that is GC-rich (80% GC), which is not uncommon for growth factor receptors (Kozak (1991) J. Cell Biol., 115, 887-903). The 3' untranslated sequence comprises 193 nucleotides and ends with a poly-A tail. No bona fide poly-A addition signal is found, but there is a sequence (AATACA), 17-22 nucleotides upstream of the poly-A tail, which may serve as a poly-A addition signal.

ALK-2 cDNA was cloned by screening an amplified oligo (dT) primed human placenta cDNA library with a radiolabelled insert derived from the PCR product 11.2. Two clones, termed HP53 and HP64, with insert sizes of 2.7 kb and 2.4 kb respectively, were identified and their sequences were determined. No sequence difference in the overlapping clones was found, suggesting they are both derived from transcripts of the same gene.

20 The cDNA for human ALK-3 was cloned by initially
screening an oligo (dT) primed human foreskin fibroblast
cDNA library with an oligonucleotide (SEQ ID No. 23)
derived from the PCR product 5.2. One positive cDNA clone
with an insert size of 3 kb, termed ON11, was identified.
25 However, upon partial sequencing, it appeared that this
clone was incomplete; it encodes only part of the kinase
domain and lacks the extracellular domain. The most 5'
sequence of ON11, a 540 nucleotide XbaI restriction
fragment encoding a truncated kinase domain, was
30 subsequently used to probe a random primed fibroblast cDNA
library from which one cDNA clone with an insert size of 3
kb, termed ONF5, was isolated (SEQ ID No. 5). Sequence
analysis of ONF5 revealed a sequence of 2932 nucleotides
without a poly-A tail, suggesting that this clone was
35 derived by internal priming. The longest open reading
frame codes for a protein of 532 amino-acids. ^{SEQ ID NO: 6} The first
ATG codon which is compatible with Kozak's consensus

sequence (Kozak, supra), is at 310-312 nucleotides and is preceded by an in-frame stop codon. The 5' and 3' untranslated sequences are 309 and 1027 nucleotides long, respectively.

5 ALK-4 cDNA was identified by screening a human oligo (dT) primed human erythroleukemia cDNA library with the radiolabelled insert of the PCR product 11.1 as a probe. One cDNA clone, termed 11H8, was identified with an insert size of 2 kb (SEQ ID No. 7). An open reading frame was
10 found encoding a protein sequence of 383 amino-acids^{SEQ ID No. 8} encoding a truncated extracellular domain with high similarity to receptor serine/threonine kinases. The 3' untranslated sequence is 818 nucleotides and does not contain a poly-A tail, suggesting that the cDNA was
15 internally primed. cDNA encoding the complete extracellular domain (nucleotides 1-366) was obtained from HEL cells by RT-PCR with 5' primer (SEQ ID No. 24) derived in part from sequence at translation start site of SKR-2 (a cDNA sequence deposited in GenBank data base, accession
20 number L10125, that is identical in part to ALK-4) and 3' primer (SEQ ID No. 25) derived from 11H8 cDNA clone.

ALK-5 was identified by screening the random primed HEL cell λ gt 10 cDNA library with the PCR product 11.1 as a probe. This yielded one positive clone termed EMBLA
25 (insert size of 5.3 kb with 2 internal EcoRI sites). Nucleotide sequencing revealed an open reading frame of 1509 bp, coding for 503 amino-acids. The open reading frame was flanked by a 5' untranslated sequence of 76 bp, and a 3' untranslated sequence of 3.7 kb which was not
30 completely sequenced. The nucleotide and deduced amino-acid sequences of ALK-5 are shown in SEQ ID Nos. 9 and 10. In the 5' part of the open reading frame, only one ATG codon was found; this codon fulfils the rules of translation initiation (Kozak, supra). An in-frame stop
35 codon was found at nucleotides (-54)-(-52) in the 5' untranslated region. The predicted ATG start codon is followed by a stretch of hydrophobic amino-acid residues

5

10

35

To summarise, clones HP22, HP57, ONF1, ONF3, ONF4 and HP29 encode the same gene, ALK-1. Clone AM6 encodes mouse ALK-1. HP53, HP64 and HP84 encode the same gene, ALK-2. ONF5, ONF2 and ON11 encode the same gene ALK-3. ME-7 and ME-D encode the mouse counterpart of human ALK-3. 11H8 encodes a different gene ALK-4, whilst 8a1 encodes the mouse equivalent. EMBLA encodes ALK-5, and ME-6 encodes ALK-6.

5

10

15

20

25

30

35

The catalytic domains of kinases can be divided into 12 subdomains with stretches of conserved amino-acid

residues. The key motifs are found in serine/threonine kinase receptors suggesting that they are functional kinases. The consensus sequence for the binding of ATP (Gly-X-Gly-X-X-Gly¹ in subdomain I followed by a Lys residue further downstream in subdomain II) is found in all the ALKs.

The kinase domains of daf-1, ActR-II, and ALKs show approximately equal sequence similarity with tyrosine and serine/threonine protein kinases. However analysis of the amino-acid sequences in subdomains VI and VIII, which are the most useful to distinguish a specificity for phosphorylation of tyrosine residues versus serine/threonine residues (Hanks et al (1988) Science 241 42-52) indicates that these kinases are serine/threonine kinases; refer to Table 2.

0903917-04198

TABLE 2

KINASE	SUBDOMAINS	
	VIB	VIII
C Serine/threonine kinase consensus	DLKPEN SEQ ID NO: 38	G (T/S) XX (Y/F) X _A SEQ ID NO: 43
5 Tyrosine kinase consensus	DLAARN SEQ ID NO: 39	XP (I/V) (K/R) W (T/M) SEQ ID NO: 44 A
Act R-II	DIKSKN SEQ ID NO: 40	GTRRYM _A SEQ ID NO: 45
Act R-IIB	DFKSKN SEQ ID NO: 41	GTRRYM _A SEQ ID NO: 45
TBR-II	DLKSSN SEQ ID NO: 42	GTARYM _A SEQ ID NO: 46
ALK-I	DFKSRN SEQ ID NO: 27	GTKRYM _A SEQ ID NO: 29
10 ALK -2, -3, -4, -5, & -6	DLKSKN SEQ ID NO: 28	GTKRYM _A SEQ ID NO: 29

The sequence motifs DLKSKN (Subdomain VIB) and GTKRYM (Subdomain VIII), that are found in most of the serine/threonine kinase receptors, agree well with the consensus sequences for all protein serine/threonine kinase receptors in these regions. In addition, these receptors, except for ALK-1, do not have a tyrosine residue surrounded by acidic residues between subdomains VII and VIII, which is common for tyrosine kinases. A unique characteristic of the members of the ALK serine/threonine kinase receptor family is the presence of two short inserts in the kinase

domain between subdomains VIA and VIB and between subdomains X and XI. In the intracellular domain, these regions, together with the juxtamembrane part and C-terminal tail, are the most divergent between family members (see Figures 3 and 4). Based on the sequence similarity with the type II receptors for TGF- β and activin, the C termini of the kinase domains of ALKs -1 to -6 are set at Ser-495, Ser-501, Ser-527, Gln-500, Gln-498 and Ser-497, respectively.

10 mRNA Expression

The distribution of ALK-1, -2, -3, -4 was determined by Northern blot analysis. A Northern blot filter with mRNAs from different human tissues was obtained from Clontech (Palo Alto, C.A.). The filters were hybridized with ^{32}P -labelled probes at 42°C overnight in 50% formaldehyde, 5 x standard saline citrate (SSC; 1xSSC is 50mM sodium citrate, pH 7.0, 150 mM NaCl), 0.1% SDS, 50 mM sodium phosphate, 5 x Denhardt's solution and 0.1 mg/ml salmon sperm DNA. In order to minimize cross-hybridization, probes were used that did not encode part of the kinase domains, but corresponded to the highly diverged sequences of either 5' untranslated and ligand-binding regions (probes for ALK-1, -2 and -3) or 3' untranslated sequences (probe for ALK-4). The probes were labelled by random priming using the Multiprime (or Mega-prime) DNA labelling system and [α - ^{32}P] dCTP (Feinberg & Vogelstein (1983) Anal. Biochem. 132: 6-13). Unincorporated label was removed by Sephadex G-25 chromatography. Filters were washed at 65°C, twice for 30 minutes in 2.5 x SSC, 0.1% SDS and twice for 30 minutes in 0.3 x SSC, 0.1% SDS before being exposed to X-ray film. Stripping of blots was performed by incubation at 90-100°C in water for 20 minutes.

The ALK-5 mRNA size and distribution were determined by Northern blot analysis as above. An EcoRI fragment of 980bp of the full length ALK-5 cDNA clone, corresponding to the C-terminal part of the kinase domain and 3'

00039177-031398

Using the probe for ALK-1, two transcripts of 2.2 and 4.9kb were detected. The ALK-1 expression level varied strongly between different tissues, high in placenta and lung, moderate in heart, muscle and kidney, and low (to not detectable) in brain, liver and pancreas. The relative ratios between the two transcripts were similar in most tissues; in kidney, however, there was relatively more of the 4.9 kb transcript. By reprobing the blot with a probe for ALK-2, one transcript of 4.0 kb was detected with a ubiquitous expression pattern. Expression was detected in every tissue investigated and was highest in placenta and skeletal muscle. Subsequently the blot was reprobed for ALK-3. One major transcript of 4.4 kb and a minor transcript of 7.9 kb were detected. Expression was high in skeletal muscle, in which also an additional minor transcript of 10 kb was observed. Moderate levels of ALK-3 mRNA were detected in heart, placenta, kidney and pancreas, and low (to not detectable) expression was found in brain, lung and liver. The relative ratios between the different transcripts were similar in the tested tissues, the 4.4 kb transcript being the predominant one, with the exception for brain where both transcripts were expressed at a similar level. Probing the blot with ALK-4 indicated the presence of a transcript with the estimated size of 5.2 kb and revealed an ubiquitous expression pattern. The results of Northern blot analysis using the probe for ALK-5 showed that a 5.5 kb transcript is expressed in all human tissues tested, being most abundant in placenta and least abundant in brain and heart.

The distribution of mRNA for mouse ALK-3 and -6 in various mouse tissues was also determined by Northern blot analysis. A multiple mouse tissue blot was obtained from Clontech, Palo Alto, California, U.S.A. The filter was hybridized as described above with probes for mouse ALK-3

and ALK-6. The EcoRI-PstI restriction fragment, corresponding to nucleotides 79-1100 of ALK-3, and the SacI-HpaI fragment, corresponding to nucleotides 57-720 of ALK-6, were used as probes. The filter was washed at 65°C
 5 twice for 30 minutes in 2.5 x SSC, 0.1% SDS and twice for 30 minutes with 0.3 x SSC, 0.1% SDS and then subjected to autoradiography.

Using the probe for mouse ALK-3, a 1.1 kb transcript was found only in spleen. By reprobing the blot with the
 10 ALK-6 specific probe, a transcript of 7.2 kb was found in brain and a weak signal was also seen in lung. No other signal was seen in the other tissues tested, i.e. heart, liver, skeletal muscle, kidney and testis.

All detected transcript sizes were different, and thus
 15 no cross-reaction between mRNAs for the different ALKs was observed when the specific probes were used. This suggests that the multiple transcripts of ALK-1 and ALK-3 are coded from the same gene. The mechanism for generation of the different transcripts is unknown at present; they may be
 20 formed by alternative mRNA splicing, differential polyadenylation, use of different promoters, or by a combination of these events. Differences in mRNA splicing in the regions coding for the extracellular domains may lead to the synthesis of receptors with different
 25 affinities for ligands, as was shown for mActR-IIB (Attisano *et al* (1992) Cell 68, 97-108) or to the production of soluble binding protein.

The above experiments describe the isolation of nucleic acid sequences coding for new family of human
 30 receptor kinases. The cDNA for ALK-5 was then used to determine the encoded protein size and binding properties.
Properties of the ALKs cDNA Encoded Proteins

To study the properties of the proteins encoded by the different ALK cDNAs, the cDNA for each ALK was subcloned
 35 into a eukaryotic expression vector and transfected into various cell types and then subjected to immunoprecipitation using a rabbit antiserum raised against

0003917.031398
 RECEIVED 7/26/98

5 used:

ALK-2 151-172

ALK-4 153-171

ALK-6 151-168

The peptides were synthesized with an Applied Biosystems 430A Peptide Synthesizer using t-butoxycarbonyl

Transient transfection of the ALK-5 cDNA

COS-1 cells (American Type Culture Collection) and the R mutant of Mv1Lu cells (for references, see below) were cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum (FBS) and 100 units/ml penicillin and 50 μ g/ml streptomycin in 5% CO₂ atmosphere at 37°C. The ALK-5 cDNA (nucleotides (-76) - 2232), which includes the complete coding region, was cloned in the pSV7d vector (Truett et al, (1985) DNA 4, 333-349), and used for transfection. Transfection into COS-1 cells was performed by the calcium phosphate precipitation method (Wigler et al (1979) Cell 16, 777-785). Briefly, cells were seeded into 6-well cell culture plates at a density of 5×10^5 cells/well, and transfected the following day with 10 μ g of recombinant plasmid. After overnight incubation, cells were washed three times with a buffer containing 25 mM Tris-HCl, pH 7.4, 138 mM NaCl, 5 mM KCl, 0.7 mM CaCl₂, 0.5

mM $MgCl_2$ and 0.6 mM Na_2HPO_4 , and then incubated with
 Dulbecco's modified Eagle's medium containing FBS and
 antibiotics. Two days after transfection, the cells were
 metabolically labelled by incubating the cells for 6 hours
 5 in methionine and cysteine-free MCDB 104 medium with 150
 $\mu Ci/ml$ of [^{35}S]-methionine and [^{35}S]-cysteine (in *vivo*
 labelling mix; Amersham). After labelling, the cells were
 washed with 150 mM NaCl, 25 mM Tris-HCl, pH 7.4, and then
 solubilized with a buffer containing 20mM Tris-HCl, pH 7.4,
 10 150 mM NaCl, 10 mM EDTA, 1% Triton X-100, 1% deoxycholate,
 1.5% Trasylol (Bayer) and 1 mM phenylmethylsulfonylfluoride
 (PMSF; Sigma). After 15 minutes on ice, the cell lysates
 were pelleted by centrifugation, and the supernatants were
 then incubated with 7 μl of preimmune serum for 1.5 hours
 15 at 4°C. Samples were then given 50 μl of protein A-
 Sepharose (Pharmacia-LKB) slurry (50% packed beads in 150
 mM NaCl, 20 mM Tris-HCl, pH 7.4, 0.2% Triton X100) and
 incubated for 45 minutes at 4°C. The beads were spun down
 by centrifugation, and the supernatants (1 ml) were then
 20 incubated with either 7 μl of preimmune serum or the VPN
 antiserum for 1.5 hours at 4°C. For blocking, 10 μg of
 peptide was added together with the antiserum. Immune
 complexes were then given 50 μl of protein A-Sepharose
 (Pharmacia - LKB) slurry (50% packed beads in 150 mM NaCl,
 25 20mM Tris-HCl, pH 7.4, 0.2% Triton X-100) and incubated for
 45 minutes at 4°C. The beads were spun down and washed
 four times with a washing buffer (20 mM Tris-HCl, pH 7.4,
 500 mM NaCl, 1% Triton X-100, 1% deoxycholate and 0.2%
 SDS), followed by one wash in distilled water. The immune
 30 complexes were eluted by boiling for 5 minutes in the SDS-
 sample buffer (100 mM Tris-HCl, pH 8.8, 0.01% bromophenol
 blue, 36% glycerol, 4% SDS) in the presence of 10 mM DTT,
 and analyzed by SDS-gel electrophoresis using 7-15%
 polyacrylamide gels (Blobel and Dobberstein, (1975) J.Cell
 35 Biol. 67, 835-851). Gels were fixed, incubated with
 Amplify (Amersham) for 20 minutes, and subjected to
 fluorography. A component of 53Da was seen. This

component was not seen when preimmune serum was used, or when 10 µg blocking peptide was added together with the antiserum. Moreover, it was not detectable in samples derived from untransfected COS-1 cells using either preimmune serum or the antiserum.

Digestion with Endoglycosidase F

Samples immunoprecipitated with the VPN antisera obtained as described above were incubated with 0.5 U of endoglycosidase F (Boehringer Mannheim Biochemica) in a buffer containing 100 mM sodium phosphate, pH 6.1, 50 mM EDTA, 1% Triton X-100, 0.1% SDS and 1% β-mercaptoethanol at 37°C for 24 hours. Samples were eluted by boiling for 5 minutes in the SDS-sample buffer, and analyzed by SDS-polyacrylamide gel electrophoresis as described above. Hydrolysis of N-linked carbohydrates by endoglycosidase F shifted the 53 kDa band to 51 kDa. The extracellular domain of ALK-5 contains one potential acceptor site for N-glycosylation and the size of the deglycosylated protein is close to the predicted size of the core protein.

Establishment of PAE Cell Lines Expressing ALK-5

In order to investigate whether the ALK-5 cDNA encodes a receptor for TGF-β, porcine aortic endothelial (PAE) cells were transfected with an expression vector containing the ALK-5 cDNA, and analyzed for the binding of ¹²⁵I-TGF-β1.

PAE cells were cultured in Ham's F-12 medium supplemented with 10% FBS and antibiotics (Miyazono *et al.*, (1988) J. Biol. Chem. 263, 6407-6415). The ALK-5 cDNA was cloned into the cytomegalovirus (CMV)-based expression vector pCDNA I/NEO (Invitrogen), and transfected into PAE cells by electroporation. After 48 hours, selection was initiated by adding Geneticin (G418 sulphate; Gibco - BRL) to the culture medium at a final concentration of 0.5 mg/ml (Westermarck *et al.*, (1990) Proc. Natl. Acad. Sci. USA 87, 128-132). Several clones were obtained, and after analysis by immunoprecipitation using the VPN antiserum, one clone denoted PAE/TBR-1 was chosen and further analyzed.

Iodination of TGF- β 1, Binding and Affinity Crosslinking

Recombinant human TGF- β 1 was iodinated using the chloramine T method according to Frolik *et al.*, (1984) *J. Biol. Chem.* **259**, 10995-11000. Cross-linking experiments were performed as previously described (Ichijo *et al.*, (1990) *Exp. Cell Res.* **187**, 263-269). Briefly, cells in 6-well plates were washed with binding buffer (phosphate-buffered saline containing 0.9 mM CaCl₂, 0.49 mM MgCl₂ and 1 mg/ml bovine serum albumin (BSA)), and incubated on ice in the same buffer with ¹²⁵I-TGF- β 1 in the presence or absence of excess unlabelled TGF- β 1 for 3 hours. Cells were washed and cross-linking was done in the binding buffer without BSA together with 0.28 mM disuccinimidyl suberate (DSS; Pierce Chemical Co.) for 15 minutes on ice. The cells were harvested by the addition of 1 ml of detachment buffer (10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 10% glycerol, 0.3 mM PMSF). The cells were pelleted by centrifugation, then resuspended in 50 μ l of solubilization buffer (125 mM NaCl, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 1% Triton X-100, 0.3 mM PMSF, 1% Trasylol) and incubated for 40 minutes on ice. Cells were centrifuged again and supernatants were subjected to analysis by SDS-gel electrophoresis using 4-15% polyacrylamide gels, followed by autoradiography. ¹²⁵I-TGF- β 1 formed a 70 kDa cross-linked complex in the transfected PAE cells (PAE/TBR-I cells). The size of this complex was very similar to that of the TGF- β type I receptor complex observed at lower amounts in the untransfected cells. A concomitant increase of 94 kDa TGF- β type II receptor complex could also be observed in the PAE/TBR-I cells. Components of 150-190 kDa, which may represent crosslinked complexes between the type I and type II receptors, were also observed in the PAE/TBR-I cells.

In order to determine whether the cross-linked 70 kDa complex contained the protein encoded by the ALK-5 cDNA, the affinity cross-linking was followed by immunoprecipitation using the VPN antiserum. For this,

0903917.031398
RECEIVED 1/27/90

cells in 25 cm² flasks were used. The supernatants obtained after cross-linking were incubated with 7 μ l of preimmune serum or VPN antiserum in the presence or absence of 10 μ g of peptide for 1.5h at 4°C. Immune complexes were then added to 50 μ l of protein A-Sepharose slurry and incubated for 45 minutes at 4°C. The protein A-Sepharose beads were washed four times with the washing buffer, once with distilled water, and the samples were analyzed by SDS-gel electrophoresis using 4-15% polyacrylamide gradient gels and autoradiography. A 70 kDa cross-linked complex was precipitated by the VPN antiserum in PAE/TBR-1 cells, and a weaker band of the same size was also seen in the untransfected cells, indicating that the untransfected PAE cells contained a low amount of endogenous ALK-5. The 70 kDa complex was not observed when preimmune serum was used, or when immune serum was blocked by 10 μ g of peptide. Moreover, a coprecipitated 94 kDa component could also be observed in the PAE/TBR-I cells. The latter component is likely to represent a TGF- β type II receptor complex, since an antiserum, termed DRL, which was raised against a synthetic peptide from the C-terminal part of the TGF- β type II receptor, precipitated a 94 kDa TGF- β type II receptor complex, as well as a 70 kDa type I receptor complex from PAE/TBR-I cells.

The carbohydrate contents of ALK-5 and the TGF- β type II receptor were characterized by deglycosylation using endoglycosidase F as described above and analyzed by SDS-polyacrylamide gel electrophoresis and autoradiography. The ALK-5 cross-linked complex shifted from 70 kDa to 66 kDa, whereas that of the type II receptor shifted from 94 kDa to 82 kDa. The observed larger shift of the type II receptor band compared with that of the ALK-5 band is consistent with the deglycosylation data of the type I and type II receptors on rat liver cells reported previously (Cheifetz *et al* (1988) *J. Biol. Chem.* 263, 16984-16991), and fits well with the fact that the porcine TGF- β type II receptor has two N-glycosylation sites (Lin *et al* (1992)

Cell 68, 775-785), whereas ALK-5 has only one (see SEQ ID No. 9).

Binding of TGF- β 1 to the type I receptor is known to be abolished by transient treatment of the cells with dithiothreitol (DTT) (Cheifetz and Massague (1991) J. Biol. Chem. 266, 20767-20772; Wrana et al (1992) Cell 71, 1003-1014). When analyzed by affinity cross-linking, binding of 125 I-TGF- β 1 to ALK-5, but not to the type II receptor, was completely abolished by DTT treatment of PAE/TBR-1 cells. Affinity cross-linking followed by immunoprecipitation by the VPN antiserum showed that neither the ALK-5 nor the type II receptor complexes was precipitated after DTT treatment, indicating that the VPN antiserum reacts only with ALK-5. The data show that the VPN antiserum recognizes a TGF- β type I receptor, and that the type I and type II receptors form a heteromeric complex.

125 I-TGF- β 1 Binding & Affinity Crosslinking of Transfected COS Cells

Transient expression plasmids of ALKs -1 to -6 and TBR-II were generated by subcloning into the pSV7d expression vector or into the pcDNA I expression vector (Invitrogen). Transient transfection of COS-1 cells and iodination of TGF- β 1 were carried out as described above. Crosslinking and immunoprecipitation were performed as described for PAE cells above.

Transfection of cDNAs for ALKs into COS-1 cells did not show any appreciable binding of 125 I-TGF β 1, consistent with the observation that type I receptors do not bind TGF- β in the absence of type II receptors. When the TBR-II cDNA was co-transfected with cDNAs for the different ALKs, type I receptor-like complexes were seen, at different levels, in each case. COS-1 cells transfected with TBR-II and ALK cDNAs were analyzed by affinity crosslinking followed by immunoprecipitation using the DRL antisera or specific antisera against ALKs. Each one of the ALKs bound 125 I-TGF- β 1 and was coimmunoprecipitated with the TBR-II complex using the DRL antiserum. Comparison of the

0003917.034393
SECRET//T6E060

5

35

30 Next, it was investigated whether ALKs could restore
responsiveness to TGF- β in the R mutant of Mv1Lu cells,
which lack the ligand-binding ability of the TGF- β type I
receptor but have intact type II receptor. Wild-type Mv1Lu
cells and mutant cells were transfected with ALK cDNA and
35 were then assayed for the production of plasminogen
activator inhibitor-1 (PAI-1) which is produced as a result
of TGF- β receptor activation as described previously by

Laiho *et al* (1991) Mol. Cell Biol. 11, 972-978. Briefly, cells were added with or without 10 ng/ml of TGF- β 1 for 2 hours in serum-free MCDB 104 without methionine. Thereafter, cultures were labelled with [35 S] methionine (40 μ Ci/ml) for 2 hours. The cells were removed by washing on ice once in PBS, twice in 10 mM Tris-HCl (pH 8.0), 0.5% sodium deoxycholate, 1 mM PMSF, twice in 2 mM Tris-HCl (pH 8.0), and once in PBS. Extracellular matrix proteins were extracted by scraping cells into the SDS-sample buffer containing DTT, and analyzed by SDS-gel electrophoresis followed by fluorography using Amplify. PAI-1 can be identified as a characteristic 45kDa band (Laiho *et al* (1991) Mol. Cell Biol. 11, 972-978). Wild-type Mv1Lu cells responded to TGF- β and produced PAI-1, whereas the R mutant clone did not, even after stimulation by TGF- β 1. Transient transfection of the ALK-5 cDNA into the R mutant clone led to the production of PAI-1 in response to the stimulation by TGF- β 1, indicating that the ALK-5 cDNA encodes a functional TGF- β type I receptor. In contrast, the R mutant cells that were transfected with other ALKs did not produce PAI-1 upon the addition of TGF- β 1.

Using similar approaches as those described above for the identification of TGF- β -binding ALKs, the ability of ALKs to bind activin in the presence of ActRII was examined. COS-1 cells were co-transfected as described above. Recombinant human activin A was iodinated using the chloramine T method (Mathews and Vale (1991) Cell 65, 973-982). Transfected COS-1 cells were analysed for binding and crosslinking of 125 I-activin A in the presence or absence of excess unlabelled activin A. The crosslinked complexes were subjected to immunoprecipitation using DRL antisera or specific ALK antisera.

All ALKs appear to bind activin A in the presence of Act R-II. This is more clearly demonstrated by affinity cross-linking followed by immunoprecipitation. ALK-2 and ALK-4 bound 125 I-activin A and were coimmunoprecipitated

In order to investigate whether ALKs are physiological activin type I receptors, activin responsive cells were examined for the expression of endogenous activin type I receptors. Mv1Lu cells, as well as the R mutant, express both type I and type II receptors for activin, and the R mutant cells produce PAI-1 upon the addition of activin A. Mv1Lu cells were labeled with ¹²⁵I-activin A, cross-linked and immunoprecipitated by the antisera against ActR-II or ALKs as described above.

Similar to Mv1Lu cells, activin type I receptor complexes in PAE/Chim A cells were immunoprecipitated by the ALK-2 and ALK-4 antisera. These results show that both ALK-2 and ALK-4 serve as high affinity type I receptors for activin A in these cells.

ALK-1, ALK-3 and ALK-6 bind TGF- β 1 and activin A in the presence of their respective type II receptors, but the

functional consequences of the binding of the ligands remains to be elucidated.

The experiments described supra suggested further experiments. Specifically, it is known that TGF- β family members act as ligands in connection with specific type I and type II receptors, with resulting complexes interacting with members of the Smad family. See Heldin et al., Nature 390: 465-471 (1997), incorporated by reference. The Smad molecules are homologs of molecules found in *Drosophila* ("Mad"), and *C. elegans* (Sma), hence, the acronym "Smad". These are involved in signal transduction pathways downstream of serine/threonine kinase receptors. See Massagué et al., Trends Cell Biol. 2: 187-192 (1997). The different members of the family have different signalling roles. Smad1, for example, as well as Smad 2 and 3, and perhaps Smad 5, become phosphorylated via specific type 1 serine/threonine kinase receptors, and act in pathway restricted fashion. For example, *Xenopus* Mad1 induces ventral mesoderm, in the presence of BMP. The human Smad1 has been shown to have ventralizing activity. See Liu et al., Nature 381: 620-623 (1996); Kretzschmer et al., Genes Dev 11: 984-995 (1997). There is also some evidence that TGF- β phosphorylates Smad1. See Lechleider et al., J. Biol. Chem. 271: 17617-17620 (1996); Yingling et al., Proc. Natl. Acad. Sci. USA 93: 8940-8944 (1996). Given what was known regarding this complex signalling pathway, the role of ALK-1 was studied.

COS-7 cells, which do not express ALK-1, were transfected with cDNA encoding tagged ALK-1. The tag was hemagglutinin (hereafter "HA"), and a commercially available lipid containing transfecting agent was used. In parallel experiments, porcine aortic endothelial (PAE) cells were also used, because these cells express TGF β type II receptors, and ALK-5, but not ALK-1. Hence, PAE cells were either transfected, or not. Transfection protocols are given, supra.

The cells were then contacted with 125 I labelled TGF- β 1, and were then contacted with ALK-1 specific antisera, to ascertain

SECRET 2/25/90

whether cross linking had occurred. See the experiments, supra, as well as ten Dijke et al., Science 264: 101-104 (1994), incorporated by reference. Antisera to ALK-5 were also used.

5 The results indicated that the ALK-1 antiserum immunoprecipitated complexes of the appropriate size from the transfected COS-7 and PAE cells, but not those which were not transfected, thereby establishing that ALK-1 is a receptor for TGF- β .

10 This was confirmed in experiments on human umbilical vein endothelial cells (HUVEC). These cells are known to express ALK-1 endogenously, as well as ALK-5. The ALK-5 antiserum and the ALK-1 antiserum both immunoprecipitated type I and type II receptor cross linked complexes. The ALK-1 antiserum immunoprecipitated band migrated slightly more slowly than the band immunoprecipitated by the ALK-5 antiserum. This is in agreement with the difference in size of ALK-1 and ALK-5, and it indicates that both ALK-1 and ALK-5 bind TGF- β 1 in HUVECS.

15 Further, it shows that ALK-1 acts as a co-called "type I" TGF- β receptor in an endogenous, physiological setting.

20 Once it was determined that TGF- β 1 and ALK-1 interact, studies were carried out to determine whether or not activation of ALK-1 resulted in phosphorylation of Smads. To test this, COS-7 cells were transfected in the same manner described supra with either Flag tagged Smad1 or Flag tagged Smad2 together with either a constitutively active form of ALK-1, or a constitutively active form of ALK-5. Specifically, the variant of ALK-1 is Q201D, and that of ALK-5 is T204D. Constitutively active ALK-1 was used to avoid the need for an additional transfection step. To elaborate, it is known that for the TGF- β pathway to function adequately, a complex of two, type I receptors, and two, type II receptors must interact, so as to activate the receptors. Constitutively active receptors, such as what was used herein, do not require the presence of the type II receptor to function. See Wieser et al., EMBO J 14: 2199-2208 (1995). In order to determine if the

25
30

resulting transfected cells produced phosphorylated Smads, Smads were determined using a Flag specific antibody, which precipitated them, and phosphorylation was determined using the antiphosphoserine antibody of Nishimura et al., J. Biol. Chem. 273: 1872-1879 (1998). It was determined, when the data were analyzed, that Smad1 was phosphorylated following interaction with activated ALK-1, but not following interaction of TGF- β and ALK-5. Conversely, the interaction of TGF- β and ALK-5 led to phosphorylation of Smad 2, but not Smad 1. This supports a conclusion that ALK-1 transduces signal in a manner similar to BMPs.

Additional experiments were then carried out to study the interaction of ALK-1 with Smad-1. Specifically, COS-7 cells were transfected with cDNA which encoded the wild type form of the TGF β type II receptor (TBR-II), a kinase inactive form of ALK-1, and Flag tagged Smad-1. Kinase inactive ALK-1 was used, because the interaction of Smad-1 and receptors is known to be transient, as once Smads are phosphorylated they dissociate from the type I receptor. See Marcias-Silva et al., Cell 87: 1215-1224 (1996); Nakao et al., EMBO J 16: 5353-5362 (1997). Affinity cross-linking, using ^{125}I -TGF- β 1, and immunoprecipitation with Flag antibody was carried out, as discussed supra. The expression of ALK-1 was determined using anti-HA antibody, since the vector used to express ALK-1 effectively tagged it with HA.

The immunoprecipitating of Smad1 resulted in coprecipitation of a cross linked TBR-II/ALK-1 complex, suggesting a direct association of Smad1 with ALK-1.

These examples show that one can identify molecules which inhibit, or enhance expression of a gene whose expression is regulated by phosphorylated Smad1. To elaborate, as ALK-1 has been identified as a key constituent of the pathway by which Smad1 is phosphorylated, one can contact cells which express both Smad1 and ALK-1 with a substance of interest, and then determine if the Smad1 becomes phosphorylated. The cells can be those which inherently

express both ALK-1 and Smad1, or which have been transformed or transfected with DNA encoding one or both of these. One can determine the phosphorylation via, e.g., the use of anti phosphorylated serine antibodies, as discussed supra. In an especially preferred embodiment, the assay can be carried out using TGF- β , as a competing agent. The TGF- β , as has been shown, does bind to ALK-1, leading to phosphorylation of Smad1. Hence, by determining a value with TGF- β alone, one can then compare a value determined with amounts of the substance to be tested, in the presence of TGF- β . Changes in phosphorylation levels can thus be attributed to the test substance.

In this type of system, it must be kept in mind that both type I receptors and type II receptors must be present; however, as indicated, supra, one can eliminate the requirement for a type II receptor by utilizing a constitutively active form of ALK-1, such as the form described supra. Additional approaches to inhibiting this system will be clear to the skilled artisan. For example, since it is known that there is interaction between Smad1 and the ALK-1 receptor, one can test for inhibition via the use of small molecules which inhibit the receptor/Smad interaction. Heldin et al., supra, mention Smad6 and Smad7 as Smad1 inhibitors, albeit in the context of a different system. Hence one can test for inhibition, or inhibit the interaction, via adding a molecule to be tested or for actual inhibition to a cell, wherein the molecule is internalized by the cell, followed by assaying for phosphorylation, via a method such as is discussed supra.

In a similar way, one can assay for inhibitors of type I/type II receptor interaction, by testing the molecule of interest in a system which includes both receptors, and then assaying for phosphorylation.

Conversely, activators or agonists can also be tested for, or utilized, following the same type of procedures.

Via using any of these systems, one can identify any gene or genes which are activated by phosphorylated Smad1. To elaborate,

the art is very familiar with systems of expression analysis, such as differential display PCR, subtraction hybridization, and other systems which combine driver and testes populations of nucleic acids, whereby transcripts which are expressed or not expressed can be identified. By simply using an activator/inhibitor of the system disclosed herein, on a first sample, and a second sample where none is used, one can then carry out analysis of transcript, thereby determining the transcripts of interest.

Also a part of the invention is the regulation of phosphorylation of Smad-1, with inhibitors, such as antibodies against the extracellular domain of ALK-1 or TGF- β , or enhancers, such as TGF- β itself, or those portions of the TGF- β molecule which are necessary for binding. Indeed, by appropriate truncation, one can also determine what portions of ALK-1 are required for phosphorylation of Smad1 to take place.

The invention has been described by way of example only, without restriction of its scope. The invention is defined by the subject matter herein, including the claims that follow the immediately following full Sequence Listings.

09099177 031398
RECEIVED 2/27/95

(1) GENERAL INFORMATION:

- (i) APPLICANTS: Miyazono, Kohei; Takeshe Imamura; ten Dijke, Peter
- (ii) TITLE OF INVENTION: Isolated ALK-1 Protein, Nucleic Acids Encoding It, And Uses Thereof
- (iii) NUMBER OF SEQUENCES: 29
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Felfe & Lynch
(B) STREET: 805 Third Avenue
(C) CITY: New York City
(D) STATE: New York
(F) ZIP: 10022
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage
(B) COMPUTER: IBM
(C) OPERATING SYSTEM: PC-DOS
(D) SOFTWARE: Wordperfect
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/436,265
(B) FILING DATE: 30-October-1995
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/GB93/02367
(B) FILING DATE: 17-November-1993
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 9224057.1
(B) FILING DATE: 17-November-1992
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 9304677.9
(B) FILING DATE: 8-March-1993
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 9304680.3
(B) FILING DATE: 8-March-1993
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 9311047.6
(B) FILING DATE: 28-May-1993

RECEIVED 11/16/95

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: 9313763.6
- (B) FILING DATE: 2-July-1993

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: 9136099.2
- (B) FILING DATE: 3-August-1993

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: 9321344.5
- (B) FILING DATE: 15-October-1993

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Hanson, Norman D.
- (B) REGISTRATION NUMBER: 37,003
- (C) REFERENCE/DOCKET NUMBER: LUD 5539

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (212) 688-9200
- (B) TELEFAX: (212) 838-3884

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1984 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 283..1791

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```

AGGAAACGGT TTATTAGGAG GGAGTGGTGG AGCTGGGCCA GGCAGGAAGA CGCTGGAATA      60
AGAAACATTT TTGCTCCAGC CCCCATCCCA GTCCCGGGAG GCTGCCGCGC CAGCTGCGCC      120
GAGCGAGCCC CTCCCCGGCT CCAGCCCGGT CCGGGGCCGC GCCGGACCCC AGCCCGCCGT      180
CCAGCGCTGG CGGTGCAACT GCGGCCGCGC GGTGGAGGGG AGGTGGCCCC GGTCCGCCGA      240
AGGCTAGCGC CCCGCCACCC GCAGAGCGGG CCCAGAGGGA CC ATG ACC TTG GGC      294
Met Thr Leu Gly

```


TCC Ser 5	CCC Pro	AGG Arg	AAA Lys	GGC Gly	CTT Leu 10	CTG Leu	ATG Met	CTG Leu	CTG Leu	ATG Met 15	GCC Ala	TTG Leu	GTG Val	ACC Thr	CAG Gln 20	342
GGA Gly	GAC Asp	CCT Pro	GTG Val	AAG Lys 25	CCG Pro	TCT Ser	CGG Arg	GGC Gly	CCG Pro 30	CTG Leu	GTG Val	ACC Thr	TGC Cys	ACG Thr 35	TGT Cys	390
GAG Glu	AGC Ser	CCA Pro	CAT His 40	TGC Cys	AAG Lys	GGG Gly	CCT Pro	ACC Thr 45	TGC Cys	CGG Arg	GGG Gly	GCC Ala	TGG Trp 50	TGC Cys	ACA Thr	438
GTA Val	GTG Val	CTG Leu 55	GTG Val	CGG Arg	GAG Glu	GAG Glu	GGG Gly 60	AGG Arg	CAC His	CCC Pro	CAG Gln	GAA Glu 65	CAT His	CGG Arg	GGC Gly	486
TGC Cys	GGG Gly 70	AAC Asn	TTG Leu	CAC His	AGG Arg	GAG Glu 75	CTC Leu	TGC Cys	AGG Arg	GGG Gly	CGC Arg 80	CCC Pro	ACC Thr	GAG Glu	TTC Phe	534
GTC Val 85	AAC Asn	CAC His	TAC Tyr	TGC Cys	TGC Cys 90	GAC Asp	AGC Ser	CAC His	CTC Leu	TGC Cys 95	AAC Asn	CAC His	AAC Asn	GTG Val	TCC Ser 100	582
CTG Leu	GTG Val	CTG Leu	GAG Glu	GCC Ala 105	ACC Thr	CAA Gln	CCT Pro	CCT Pro	TCG Ser 110	GAG Glu	CAG Gln	CCG Pro	GGA Gly	ACA Thr 115	GAT Asp	630
GGC Gly	CAG Gln	CTG Leu	GCC Ala 120	CTG Leu	ATC Ile	CTG Leu	GGC Gly	CCC Pro 125	GTG Val	CTG Leu	GCC Ala	TTG Leu	CTG Leu 130	GCC Ala	CTG Leu	678
GTG Val	GCC Ala	CTG Leu 135	GGT Gly	GTC Val	CTG Leu	GGC Gly	CTG Leu 140	TGG Trp	CAT His	GTC Val	CGA Arg	CGG Arg 145	AGG Arg	CAG Gln	GAG Glu	726
AAG Lys	CAG Gln 150	CGT Arg	GGC Gly	CTG Leu	CAC His	AGC Ser 155	GAG Glu	CTG Leu	GGA Gly	GAG Glu	TCC Ser 160	AGT Ser	CTC Leu	ATC Ile	CTG Leu	774
AAA Lys 165	GCA Ala	TCT Ser	GAG Glu	CAG Gln	GGC Gly 170	GAC Asp	ACG Thr	ATG Met	TTG Leu	GGG Gly 175	GAC Asp	CTC Leu	CTG Leu	GAC Asp	AGT Ser 180	822
GAC Asp	TGC Cys	ACC Thr	ACA Thr	GGG Gly 185	AGT Ser	GGC Gly	TCA Ser	GGG Gly	CTC Leu 190	CCC Pro	TTC Phe	CTG Leu	GTG Val	CAG Gln 195	AGG Arg	870
ACA Thr	GTG Val	GCA Ala	CGG Arg 200	CAG Gln	GTT Val	GCC Ala	TTG Leu	GTG Val 205	GAG Glu	TGT Cys	GTG Val	GGA Gly 210	AAA Lys	GGC Gly	CGC Arg	918

TAT Tyr	GGC Gly	GAA Glu 215	GTG Val	TGG Trp	CGG Arg	GGC Gly	TTG Leu 220	TGG Trp	CAC His	GGT Gly	GAG Glu	AGT Ser 225	GTG Val	GCC Ala	GTC Val	966
AAG Lys	ATC Ile 230	TTC Phe	TCC Ser	TCG Ser	AGG Arg	GAT Asp 235	GAA Glu	CAG Gln	TCC Ser	TGG Trp	TTC Phe 240	CGG Arg	GAG Glu	ACT Thr	GAG Glu	1014
ATC Ile 245	TAT Tyr	AAC Asn	ACA Thr	GTA Val 250	TTG Leu	CTC Leu	AGA Arg	CAC His	GAC Asp	AAC Asn 255	ATC Ile	CTA Leu	GGC Gly	TTC Phe	ATC Ile 260	1062
GCC Ala	TCA Ser	GAC Asp	ATG Met	ACC Thr 265	TCC Ser	CGC Arg	AAC Asn	TCG Ser	AGC Ser 270	ACG Thr	CAG Gln	CTG Leu	TGG Trp	CTC Leu 275	ATC Ile	1110
ACG Thr	CAC His	TAC Tyr	CAC His 280	GAG Glu	CAC His	GGC Gly	TCC Ser	CTC Leu 285	TAC Tyr	GAC Asp	TTT Phe	CTG Leu	CAG Gln 290	AGA Arg	CAG Gln	1158
ACG Thr	CTG Leu	GAG Glu 295	CCC Pro	CAT His	CTG Leu	GCT Ala	CTG Leu 300	AGG Arg	CTA Leu	GCT Ala	GTG Val	TCC Ser 305	GCG Ala	GCA Ala	TGC Cys	1206
GGC Gly	CTG Leu 310	GCG Ala	CAC His	CTG Leu	CAC His	GTG Val 315	GAG Glu	ATC Ile	TTC Phe	GGT Gly	ACA Thr 320	CAG Gln	GGC Gly	AAA Lys	CCA Pro	1254
GCC Ala 325	ATT Ile	GCC Ala	CAC His	CGC Arg	GAC Asp 330	TTC Phe	AAG Lys	AGC Ser	CGC Arg	AAT Asn 335	GTG Val	CTG Leu	GTC Val	AAG Lys	AGC Ser 340	1302
AAC Asn	CTG Leu	CAG Gln	TGT Cys	TGC Cys 345	ATC Ile	GCC Ala	GAC Asp	CTG Leu	GGC Gly 350	CTG Leu	GCT Ala	GTG Val	ATG Met	CAC His 355	TCA Ser	1350
CAG Gln	GGC Gly	AGC Ser	GAT Asp 360	TAC Tyr	CTG Leu	GAC Asp	ATC Ile	GGC Gly 365	AAC Asn	AAC Asn	CCG Pro	AGA Arg	GTG Val 370	GGC Gly	ACC Thr	1398
AAG Lys	CGG Arg	TAC Tyr 375	ATG Met	GCA Ala	CCC Pro	GAG Glu	GTG Val 380	CTG Leu	GAC Asp	GAG Glu	CAG Gln	ATC Ile 385	CGC Arg	ACG Thr	GAC Asp	1446
TGC Cys	TTT Phe 390	GAG Glu	TCC Ser	TAC Tyr	AAG Lys	TGG Trp 395	ACT Thr	GAC Asp	ATC Ile	TGG Trp	GCC Ala 400	TTT Phe	GGC Gly	CTG Leu	GTG Val	1494
CTG Leu 405	TGG Trp	GAG Glu	ATT Ile	GCC Ala	CGC Arg 410	CGG Arg	ACC Thr	ATC Ile	GTG Val	AAT Asn 415	GGC Gly	ATC Ile	GTG Val	GAG Glu	GAC Asp 420	1542

TAT AGA CCA CCC TTC TAT GAT GTG GTG CCC AAT GAC CCC AGC TTT GAG 1590
 Tyr Arg Pro Pro Phe Tyr Asp Val Val Pro Asn Asp Pro Ser Phe Glu
 425 430 435
 GAC ATG AAG AAG GTG GTG TGT GTG GAT CAG CAG ACC CCC ACC ATC CCT 1638
 Asp Met Lys Lys Val Val Cys Val Asp Gln Gln Thr Pro Thr Ile Pro
 440 445 450
 AAC CGG CTG GCT GCA GAC CCG GTC CTC TCA GGC CTA GCT CAG ATG ATG 1686
 Asn Arg Leu Ala Ala Asp Pro Val Leu Ser Gly Leu Ala Gln Met Met
 455 460 465
 CGG GAG TGC TGG TAC CCA AAC CCC TCT GCC CGA CTC ACC GCG CTG CGG 1734
 Arg Glu Cys Trp Tyr Pro Asn Pro Ser Ala Arg Leu Thr Ala Leu Arg
 470 475 480
 ATC AAG AAG ACA CTA CAA AAA ATT AGC AAC AGT CCA GAG AAG CCT AAA 1782
 Ile Lys Lys Thr Leu Gln Lys Ile Ser Asn Ser Pro Glu Lys Pro Lys
 485 490 495 500
 GTG ATT CAA TAGCCAGGA GCACCTGATT CCTTTCTGCC TGCAGGGGGC 1831
 Val Ile Gln
 TGGGGGGGTG GGGGGCAGTG GATGGTGCCC TATCTGGGTA GAGGTAGTGT GAGTGTGGTG 1891
 TGTGCTGGGG ATGGGCAGCT GCGCCTGCCT GCTCGGCCCC CAGCCCACCC AGCCAAAAAT 1951
 ACAGCTGGGC TGAAACCTGA AAAAAAAAAA AAA 1984

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 503 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Thr Leu Gly Ser Pro Arg Lys Gly Leu Leu Met Leu Leu Met Ala
 1 5 10 15
 Leu Val Thr Gln Gly Asp Pro Val Lys Pro Ser Arg Gly Pro Leu Val
 20 25 30
 Thr Cys Thr Cys Glu Ser Pro His Cys Lys Gly Pro Thr Cys Arg Gly
 35 40 45
 Ala Trp Cys Thr Val Val Leu Val Arg Glu Glu Gly Arg His Pro Gln
 50 55 60

"SEQUENCE"

Glu His Arg Gly Cys Gly Asn Leu His Arg Glu Leu Cys Arg Gly Arg
 65 70 75 80
 Pro Thr Glu Phe Val Asn His Tyr Cys Cys Asp Ser His Leu Cys Asn
 85 90 95
 His Asn Val Ser Leu Val Leu Glu Ala Thr Gln Pro Pro Ser Glu Gln
 100 105 110
 Pro Gly Thr Asp Gly Gln Leu Ala Leu Ile Leu Gly Pro Val Leu Ala
 115 120 125
 Leu Leu Ala Leu Val Ala Leu Gly Val Leu Gly Leu Trp His Val Arg
 130 135 140
 Arg Arg Gln Glu Lys Gln Arg Gly Leu His Ser Glu Leu Gly Glu Ser
 145 150 155 160
 Ser Leu Ile Leu Lys Ala Ser Glu Gln Gly Asp Thr Met Leu Gly Asp
 165 170 175
 Leu Leu Asp Ser Asp Cys Thr Thr Gly Ser Gly Ser Gly Leu Pro Phe
 180 185 190
 Leu Val Gln Arg Thr Val Ala Arg Gln Val Ala Leu Val Glu Cys Val
 195 200 205
 Gly Lys Gly Arg Tyr Gly Glu Val Trp Arg Gly Leu Trp His Gly Glu
 210 215 220
 Ser Val Ala Val Lys Ile Phe Ser Ser Arg Asp Glu Gln Ser Trp Phe
 225 230 235 240
 Arg Glu Thr Glu Ile Tyr Asn Thr Val Leu Leu Arg His Asp Asn Ile
 245 250 255
 Leu Gly Phe Ile Ala Ser Asp Met Thr Ser Arg Asn Ser Ser Thr Gln
 260 265 270
 Leu Trp Leu Ile Thr His Tyr His Glu His Gly Ser Leu Tyr Asp Phe
 275 280 285
 Leu Gln Arg Gln Thr Leu Glu Pro His Leu Ala Leu Arg Leu Ala Val
 290 295 300
 Ser Ala Ala Cys Gly Leu Ala His Leu His Val Glu Ile Phe Gly Thr
 305 310 315 320
 Gln Gly Lys Pro Ala Ile Ala His Arg Asp Phe Lys Ser Arg Asn Val
 325 330 335

RECEIVED
 11-16-60

Leu Val Lys Ser Asn Leu Gln Cys Cys Ile Ala Asp Leu Gly Leu Ala
 340 345 350
 Val Met His Ser Gln Gly Ser Asp Tyr Leu Asp Ile Gly Asn Asn Pro
 355 360 365
 Arg Val Gly Thr Lys Arg Tyr Met Ala Pro Glu Val Leu Asp Glu Gln
 370 375 380
 Ile Arg Thr Asp Cys Phe Glu Ser Tyr Lys Trp Thr Asp Ile Trp Ala
 385 390 395 400
 Phe Gly Leu Val Leu Trp Glu Ile Ala Arg Arg Thr Ile Val Asn Gly
 405 410 415
 Ile Val Glu Asp Tyr Arg Pro Pro Phe Tyr Asp Val Val Pro Asn Asp
 420 425 430
 Pro Ser Phe Glu Asp Met Lys Lys Val Val Cys Val Asp Gln Gln Thr
 435 440 445
 Pro Thr Ile Pro Asn Arg Leu Ala Ala Asp Pro Val Leu Ser Gly Leu
 450 455 460
 Ala Gln Met Met Arg Glu Cys Trp Tyr Pro Asn Pro Ser Ala Arg Leu
 465 470 475 480
 Thr Ala Leu Arg Ile Lys Lys Thr Leu Gln Lys Ile Ser Asn Ser Pro
 485 490 495
 Glu Lys Pro Lys Val Ile Gln
 500

85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2724 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 104..1630

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CTCCGAGTAC CCCAGTGACC AGAGTGAGAG AAGCTCTGAA CGAGGGCACG CGGCTTGAAG	60
GACTGTGGGC AGATGTGACC AAGAGCCTGC ATTAAGTTGT ACA ATG GTA GAT GGA	115
Met Val Asp Gly	
1	
GTG ATG ATT CTT CCT GTG CTT ATC ATG ATT GCT CTC CCC TCC CCT AGT	163
Val Met Ile Leu Pro Val Leu Ile Met Ile Ala Leu Pro Ser Pro Ser	
5 10 15 20	
ATG GAA GAT GAG AAG CCC AAG GTC AAC CCC AAA CTC TAC ATG TGT GTG	211
Met Glu Asp Glu Lys Pro Lys Val Asn Pro Lys Leu Tyr Met Cys Val	
25 30 35	
TGT GAA GGT CTC TCC TGC GGT AAT GAG GAC CAC TGT GAA GGC CAG CAG	259
Cys Glu Gly Leu Ser Cys Gly Asn Glu Asp His Cys Glu Gly Gln Gln	
40 45 50	
TGC TTT TCC TCA CTG AGC ATC AAC GAT GGC TTC CAC GTC TAC CAG AAA	307
Cys Phe Ser Ser Leu Ser Ile Asn Asp Gly Phe His Val Tyr Gln Lys	
55 60 65	
GGC TGC TTC CAG GTT TAT GAG CAG GGA AAG ATG ACC TGT AAG ACC CCG	355
Gly Cys Phe Gln Val Tyr Glu Gln Gly Lys Met Thr Cys Lys Thr Pro	
70 75 80	
CCG TCC CCT GGC CAA GCT GTG GAG TGC TGC CAA GGG GAC TGG TGT AAC	403
Pro Ser Pro Gly Gln Ala Val Glu Cys Cys Gln Gly Asp Trp Cys Asn	
85 90 95 100	
AGG AAC ATC ACG GCC CAG CTG CCC ACT AAA GGA AAA TCC TTC CCT GGA	451
Arg Asn Ile Thr Ala Gln Leu Pro Thr Lys Gly Lys Ser Phe Pro Gly	
105 110 115	

090399
4475060

ACA Thr	CAG Gln	AAT Asn	TTC Phe 120	CAC His	TTG Leu	GAG Glu	GTT Val	GGC Gly 125	CTC Leu	ATT Ile	ATT Ile	CTC Leu	TCT Ser 130	GTA Val	GTG Val	499
TTC Phe	GCA Ala	GTA Val 135	TGT Cys	CTT Leu	TTA Leu	GCC Ala	TGC Cys 140	CTG Leu	CTG Leu	GGA Gly	GTT Val	GCT Ala 145	CTC Leu	CGA Arg	AAA Lys	547
TTT Phe 150	AAA Lys	AGG Arg	CGC Arg	AAC Asn	CAA Gln	GAA Glu 155	CGC Arg	CTC Leu	AAT Asn	CCC Pro	CGA Arg 160	GAC Asp	GTG Val	GAG Glu	TAT Tyr	595
GGC Gly 165	ACT Thr	ATC Ile	GAA Glu	GGG Gly	CTC Leu 170	ATC Ile	ACC Thr	ACC Thr	AAT Asn	GTT Val 175	GGA Gly	GAC Asp	AGC Ser	ACT Thr	TTA Leu 180	643
GCA Ala	GAT Asp	TTA Leu	TTG Leu	GAT Asp 185	CAT His	TCG Ser	TGT Cys	ACA Thr 190	TCA Ser	GGA Gly	AGT Ser	GGC Gly	TCT Ser	GGT Gly 195	CTT Leu	691
CCT Pro	TTT Phe	CTG Leu	GTA Val 200	CAA Gln	AGA Arg	ACA Thr	GTG Val	GCT Ala 205	CGC Arg	CAG Gln	ATT Ile	ACA Thr	CTG Leu 210	TTG Leu	GAG Glu	739
TGT Cys	GTC Val	GGG Gly 215	AAA Lys	GGC Gly	AGG Arg	TAT Tyr	GGT Gly 220	GAG Glu	GTG Val	TGG Trp	AGG Arg	GGC Gly 225	AGC Ser	TGG Trp	CAA Gln	787
GGG Gly 230	GAA Glu	AAT Asn	GTT Val	GCC Ala	GTG Val	AAG Lys 235	ATC Ile	TTC Phe	TCC Ser	TCC Ser	CGT Arg 240	GAT Asp	GAG Glu	AAG Lys	TCA Ser	835
TGG Trp 245	TTC Phe	AGG Arg	GAA Glu	ACG Thr	GAA Glu 250	TTG Leu	TAC Tyr	AAC Asn	ACT Thr	GTG Val 255	ATG Met	CTG Leu	AGG Arg	CAT His	GAA Glu 260	883
AAT Asn	ATC Ile	TTA Leu	GGT Gly	TTC Phe 265	ATT Ile	GCT Ala	TCA Ser	GAC Asp 270	ATG Met	ACA Thr	TCA Ser	AGA Arg	CAC His	TCC Ser 275	AGT Ser	931
ACC Thr	CAG Gln	CTG Leu	TGG Trp 280	TTA Leu	ATT Ile	ACA Thr	CAT His	TAT Tyr 285	CAT His	GAA Glu	ATG Met	GGA Gly	TCG Ser 290	TTG Leu	TAC Tyr	979
GAC Asp	TAT Tyr	CTT Leu 295	CAG Gln	CTT Leu	ACT Thr	ACT Thr	CTG Leu 300	GAT Asp	ACA Thr	GTT Val	AGC Ser	TGC Cys 305	CTT Leu	CGA Arg	ATA Ile	1027
GTG Val 310	CTG Leu	TCC Ser	ATA Ile	GCT Ala	AGT Ser	GGT Gly 315	CTT Leu	GCA Ala	CAT His	TTG Leu	CAC His 320	ATA Ile	GAG Glu	ATA Ile	TTT Phe	1075

GAAGGAAGAT TTGACGTTGT TGTCATTGTC CAGCTGGGAC CTAATGCTGG CCTGACTGGT 1710

TGTCAGAATG GAATCCATCT GTCTCCCTCC CCAAATGGCT GCTTTGACAA GGCAGACGTC 1770

GTACCCAGCC ATGTGTTGGG GAGACATCAA AACCACCCTA ACCTCGCTCG ATGACTGTGA 1830
 ACTGGGCATT TCACGAACTG TTCACACTGC AGAGACTAAT GTTGGACAGA CACTGTTGCA 1890
 AAGGTAGGGA CTGGAGGAAC ACAGAGAAAT CCTAAAAGAG ATCTGGGCAT TAAGTCAGTG 1950
 GCTTTGCATA GCTTTCACAA GTCTCCTAGA CACTCCCCAC GGGAAACTCA AGGAGGTGGT 2010
 GAATTTTAA TCAGCAATAT TGCCTGTGCT TCTCTTCTTT ATTGCACTAG GAATTCTTTG 2070
 CATTCTTAC TTGCACTGTT ACTCTTAATT TTAAAGACCC AACTTGCCAA AATGTTGGCT 2130
 GCGTACTCCA CTGGTCTGTC TTTGGATAAT AGGAATTCAA TTTGGCAAAA CAAAATGTAA 2190
 TGTCAGACTT TGCTGCATTT TACACATGTG CTGATGTTTA CAATGATGCC GAACATTAGG 2250
 AATTGTTTAT ACACAACTTT GCAAATTATT TATTACTTGT GCACTTAGTA GTTTTTACAA 2310
 AACTGCTTTG TGCATATGTT AAAGCTTATT TTTATGTGGT CTTATGATTT TATTACAGAA 2370
 ATGTTTTTAA CACTATACTC TAAATGGAC ATTTTCTTTT ATTATCAGTT AAAATCACAT 2430
 TTTAAGTGCT TCACATTTGT ATGTGTGTAG ACTGTAAGTT TTTTTCAGTT CATATGCAGA 2490
 ACGTATTTAG CCATTACCCA CGTGACACCA CCGAATATAT TATCGATTTA GAAGCAAAGA 2550
 TTTCAGTAGA ATTTTAGTCC TGAACGCTAC GGGGAAAATG CATTTTCTTC AGAATTATCC 2610
 ATTACGTGCA TTAAACTCT GCCAGAAAAA AATAACTATT TTGTTTTAAT CTACTTTTGT 2670
 TATTTAGTAG TTATTTGTAT AAATTAAATA AACTGTTTTT AAGTCAAAAA AAAA 2724

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 509 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Val Asp Gly Val Met Ile Leu Pro Val Leu Ile Met Ile Ala Leu
 1 5 10 15

Pro Ser Pro Ser Met Glu Asp Glu Lys Pro Lys Val Asn Pro Lys Leu
 20 25 30

Tyr Met Cys Val Cys Glu Gly Leu Ser Cys Gly Asn Glu Asp His Cys
 35 40 45

Glu Gly Gln Gln Cys Phe Ser Ser Leu Ser Ile Asn Asp Gly Phe His
 50 55 60
 Val Tyr Gln Lys Gly Cys Phe Gln Val Tyr Glu Gln Gly Lys Met Thr
 65 70 75 80
 Cys Lys Thr Pro Pro Ser Pro Gly Gln Ala Val Glu Cys Cys Gln Gly
 85 90 95
 Asp Trp Cys Asn Arg Asn Ile Thr Ala Gln Leu Pro Thr Lys Gly Lys
 100 105 110
 Ser Phe Pro Gly Thr Gln Asn Phe His Leu Glu Val Gly Leu Ile Ile
 115 120 125
 Leu Ser Val Val Phe Ala Val Cys Leu Leu Ala Cys Leu Leu Gly Val
 130 135 140
 Ala Leu Arg Lys Phe Lys Arg Arg Asn Gln Glu Arg Leu Asn Pro Arg
 145 150 155 160
 Asp Val Glu Tyr Gly Thr Ile Glu Gly Leu Ile Thr Thr Asn Val Gly
 165 170 175
 Asp Ser Thr Leu Ala Asp Leu Leu Asp His Ser Cys Thr Ser Gly Ser
 180 185 190
 Gly Ser Gly Leu Pro Phe Leu Val Gln Arg Thr Val Ala Arg Gln Ile
 195 200 205
 Thr Leu Leu Glu Cys Val Gly Lys Gly Arg Tyr Gly Glu Val Trp Arg
 210 215 220
 Gly Ser Trp Gln Gly Glu Asn Val Ala Val Lys Ile Phe Ser Ser Arg
 225 230 235 240
 Asp Glu Lys Ser Trp Phe Arg Glu Thr Glu Leu Tyr Asn Thr Val Met
 245 250 255
 Leu Arg His Glu Asn Ile Leu Gly Phe Ile Ala Ser Asp Met Thr Ser
 260 265 270
 Arg His Ser Ser Thr Gln Leu Trp Leu Ile Thr His Tyr His Glu Met
 275 280 285
 Gly Ser Leu Tyr Asp Tyr Leu Gln Leu Thr Thr Leu Asp Thr Val Ser
 290 295 300
 Cys Leu Arg Ile Val Leu Ser Ile Ala Ser Gly Leu Ala His Leu His
 305 310 315 320

00033177
 03398
 00033177
 03398

Ile Glu Ile Phe Gly Thr Gln Gly Lys Pro Ala Ile Ala His Arg Asp
 325 330 335
 Leu Lys Ser Lys Asn Ile Leu Val Lys Lys Asn Gly Gln Cys Cys Ile
 340 345 350
 Ala Asp Leu Gly Leu Ala Val Met His Ser Gln Ser Thr Asn Gln Leu
 355 360 365
 Asp Val Gly Asn Asn Pro Arg Val Gly Thr Lys Arg Tyr Met Ala Pro
 370 375 380
 Glu Val Leu Asp Glu Thr Ile Gln Val Asp Cys Phe Asp Ser Tyr Lys
 385 390 395 400
 Arg Val Asp Ile Trp Ala Phe Gly Leu Val Leu Trp Glu Val Ala Arg
 405 410 415
 Arg Met Val Ser Asn Gly Ile Val Glu Asp Tyr Lys Pro Pro Phe Tyr
 420 425 430
 Asp Val Val Pro Asn Asp Pro Ser Phe Glu Asp Met Arg Lys Val Val
 435 440 445
 Cys Val Asp Gln Gln Arg Pro Asn Ile Pro Asn Arg Trp Phe Ser Asp
 450 455 460
 Pro Thr Leu Thr Ser Leu Ala Lys Leu Met Lys Glu Cys Trp Tyr Gln
 465 470 475 480
 Asn Pro Ser Ala Arg Leu Thr Ala Leu Arg Ile Lys Lys Thr Leu Thr
 485 490 495
 Lys Ile Asp Asn Ser Leu Asp Lys Leu Lys Thr Asp Cys
 500 505

RECEIVED
 11-11-60

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2932 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: cDNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (v) FRAGMENT TYPE: internal
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 310..1905
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GCTCCGCGCC GAGGGCTGGA GGATGCGTTC CCTGGGGTCC GGACTTATGA AAATATGCAT 60
 CAGTTTAATA CTGTCTTGGA ATTCATGAGA TGGAAGCATA GGTCAAAGCT GTTTGGAGAA 120
 AATCAGAAGT ACAGTTTTAT CTAGCCACAT CTTGGAGGAG TCGTAAGAAA GCAGTGGGAG 180
 TTGAAGTCAT TGTCAAGTGC TTGCGATCTT TTACAAGAAA ATCTCACTGA ATGATAGTCA 240
 TTAAATTGG TGAAGTAGCA AGACCAATTA TTAAAGGTGA CAGTACACAG GAAACATTAC 300
 AATTGAACA ATG ACT CAG CTA TAC ATT TAC ATC AGA TTA TTG GGA GCC 348
 Met Thr Gln Leu Tyr Ile Tyr Ile Arg Leu Leu Gly Ala
 1 5 10
 TAT TTG TTC ATC ATT TCT CGT GTT CAA GGA CAG AAT CTG GAT AGT ATG 396
 Tyr Leu Phe Ile Ile Ser Arg Val Gln Gly Gln Asn Leu Asp Ser Met
 15 20 25
 CTT CAT GGC ACT GGG ATG AAA TCA GAC TCC GAC CAG AAA AAG TCA GAA 444
 Leu His Gly Thr Gly Met Lys Ser Asp Ser Asp Gln Lys Lys Ser Glu
 30 35 40 45
 AAT GGA GTA ACC TTA GCA CCA GAG GAT ACC TTG CCT TTT TTA AAG TGC 492
 Asn Gly Val Thr Leu Ala Pro Glu Asp Thr Leu Pro Phe Leu Lys Cys
 50 55 60
 TAT TGC TCA GGG CAC TGT CCA GAT GAT GCT ATT AAT AAC ACA TGC ATA 540
 Tyr Cys Ser Gly His Cys Pro Asp Asp Ala Ile Asn Asn Thr Cys Ile
 65 70 75
 ACT AAT GGA CAT TGC TTT GCC ATC ATA GAA GAA GAT GAC CAG GGA GAA 588
 Thr Asn Gly His Cys Phe Ala Ile Ile Glu Glu Asp Asp Gln Gly Glu
 80 85 90

ACC Thr	ACA Thr 95	TTA Leu	GCT Ala	TCA Ser	GGG Gly	TGT Cys 100	ATG Met	AAA Lys	TAT Tyr	GAA Glu	GGA Gly 105	TCT Ser	GAT Asp	TTT Phe	CAG Gln	636
TGC Cys 110	AAA Lys	GAT Asp	TCT Ser	CCA Pro	AAA Lys 115	GCC Ala	CAG Gln	CTA Leu	CGC Arg	CGG Arg 120	ACA Thr	ATA Ile	GAA Glu	TGT Cys	TGT Cys 125	684
CGG Arg	ACC Thr	AAT Asn	TTA Leu	TGT Cys 130	AAC Asn	CAG Gln	TAT Tyr	TTG Leu	CAA Gln 135	CCC Pro	ACA Thr	CTG Leu	CCC Pro	CCT Pro 140	GTT Val	732
GTC Val	ATA Ile	GGT Gly	CCG Pro 145	TTT Phe	TTT Phe	GAT Asp	GGC Gly	AGC Ser 150	ATT Ile	CGA Arg	TGG Trp	CTG Leu	GTT Val 155	TTG Leu	CTC Leu	780
ATT Ile	TCT Ser	ATG Met 160	GCT Ala	GTC Val	TGC Cys	ATA Ile 165	ATT Ile	GCT Ala	ATG Met	ATC Ile	ATC Ile	TTC Phe 170	TCC Ser	AGC Ser	TGC Cys	828
TTT Phe	TGT Cys 175	TAC Tyr	AAA Lys	CAT His	TAT Tyr	TGC Cys 180	AAG Lys	AGC Ser	ATC Ile	TCA Ser	AGC Ser 185	AGA Arg	CGT Arg	CGT Arg	TAC Tyr	876
AAT Asn 190	CGT Arg	GAT Asp	TTG Leu	GAA Glu	CAG Gln 195	GAT Asp	GAA Glu	GCA Ala	TTT Phe	ATT Ile 200	CCA Pro	GTT Val	GGA Gly	GAA Glu	TCA Ser 205	924
CTA Leu	AAA Lys	GAC Asp	CTT Leu	ATT Ile 210	GAC Asp	CAG Gln	TCA Ser	CAA Gln	AGT Ser 215	TCT Ser	GGT Gly	AGT Ser	GGG Gly	TCT Ser 220	GGA Gly	972
CTA Leu	CCT Pro	TTA Leu	TTG Leu 225	GTT Val	CAG Gln	CGA Arg	ACT Thr	ATT Ile 230	GCC Ala	AAA Lys	CAG Gln	ATT Ile	CAG Gln 235	ATG Met	GTC Val	1020
CGG Arg	CAA Gln	GTT Val 240	GGT Gly	AAA Lys	GGC Gly	CGA Arg	TAT Tyr 245	GGA Gly	GAA Glu	GTA Val	TGG Trp	ATG Met 250	GGC Gly	AAA Lys	TGG Trp	1068
CGT Arg	GGC Gly 255	GAA Glu	AAA Lys	GTG Val	GCG Ala	GTG Val 260	AAA Lys	GTA Val	TTC Phe	TTT Phe	ACC Thr 265	ACT Thr	GAA Glu	GAA Glu	GCC Ala	1116
AGC Ser 270	TGG Trp	TTT Phe	CGA Arg	GAA Glu	ACA Thr 275	GAA Glu	ATC Ile	TAC Tyr	CAA Gln	ACT Thr 280	GTG Val	CTA Leu	ATG Met	CGC Arg	CAT His 285	1164
GAA Glu	AAC Asn	ATA Ile	CTT Leu	GGT Gly 290	TTC Phe	ATA Ile	GCG Ala	GCA Ala	GAC Asp 295	ATT Ile	AAA Lys	GGT Gly	ACA Thr	GGT Gly 300	TCC Ser	1212

TGG	ACT	CAG	CTC	TAT	TTG	ATT	ACT	GAT	TAC	CAT	GAA	AAT	GGA	TCT	CTC	1260
Trp	Thr	Gln	Leu	Tyr	Leu	Ile	Thr	Asp	Tyr	His	Glu	Asn	Gly	Ser	Leu	
			305					310					315			
TAT	GAC	TTC	CTG	AAA	TGT	GCT	ACA	CTG	GAC	ACC	AGA	GCC	CTG	CTT	AAA	1308
Tyr	Asp	Phe	Leu	Lys	Cys	Ala	Thr	Leu	Asp	Thr	Arg	Ala	Leu	Leu	Lys	
		320					325					330				
TTG	GCT	TAT	TCA	GCT	GCC	TGT	GGT	CTG	TGC	CAC	CTG	CAC	ACA	GAA	ATT	1356
Leu	Ala	Tyr	Ser	Ala	Ala	Cys	Gly	Leu	Cys	His	Leu	His	Thr	Glu	Ile	
	335					340					345					
TAT	GGC	ACC	CAA	GGA	AAG	CCC	GCA	ATT	GCT	CAT	CGA	GAC	CTA	AAG	AGC	1404
Tyr	Gly	Thr	Gln	Gly	Lys	Pro	Ala	Ile	Ala	His	Arg	Asp	Leu	Lys	Ser	
350					355					360					365	
AAA	AAC	ATC	CTC	ATC	AAG	AAA	AAT	GGG	AGT	TGC	TGC	ATT	GCT	GAC	CTG	1452
Lys	Asn	Ile	Leu	Ile	Lys	Lys	Asn	Gly	Ser	Cys	Cys	Ile	Ala	Asp	Leu	
				370					375					380		
GGC	CTT	GCT	GTT	AAA	TTC	AAC	AGT	GAC	ACA	AAT	GAA	GTT	GAT	GTG	CCC	1500
Gly	Leu	Ala	Val	Lys	Phe	Asn	Ser	Asp	Thr	Asn	Glu	Val	Asp	Val	Pro	
			385					390					395			
TTG	AAT	ACC	AGG	GTG	GGC	ACC	AAA	CGC	TAC	ATG	GCT	CCC	GAA	GTG	CTG	1548
Leu	Asn	Thr	Arg	Val	Gly	Thr	Lys	Arg	Tyr	Met	Ala	Pro	Glu	Val	Leu	
		400					405					410				
GAC	GAA	AGC	CTG	AAC	AAA	AAC	CAC	TTC	CAG	CCC	TAC	ATC	ATG	GCT	GAC	1596
Asp	Glu	Ser	Leu	Asn	Lys	Asn	His	Phe	Gln	Pro	Tyr	Ile	Met	Ala	Asp	
	415					420					425					
ATC	TAC	AGC	TTC	GGC	CTA	ATC	ATT	TGG	GAG	ATG	GCT	CGT	CGT	TGT	ATC	1644
Ile	Tyr	Ser	Phe	Gly	Leu	Ile	Ile	Trp	Glu	Met	Ala	Arg	Arg	Cys	Ile	
430					435					440					445	
ACA	GGA	GGG	ATC	GTG	GAA	GAA	TAC	CAA	TTG	CCA	TAT	TAC	AAC	ATG	GTA	1692
Thr	Gly	Gly	Ile	Val	Glu	Glu	Tyr	Gln	Leu	Pro	Tyr	Tyr	Asn	Met	Val	
				450					455					460		
CCG	AGT	GAT	CCG	TCA	TAC	GAA	GAT	ATG	CGT	GAG	GTT	GTG	TGT	GTC	AAA	1740
Pro	Ser	Asp	Pro	Ser	Tyr	Glu	Asp	Met	Arg	Glu	Val	Val	Cys	Val	Lys	
			465					470					475			
CGT	TTG	CGG	CCA	ATT	GTG	TCT	AAT	CGG	TGG	AAC	AGT	GAT	GAA	TGT	CTA	1788
Arg	Leu	Arg	Pro	Ile	Val	Ser	Asn	Arg	Trp	Asn	Ser	Asp	Glu	Cys	Leu	
		480					485					490				
CGA	GCA	GTT	TTG	AAG	CTA	ATG	TCA	GAA	TGC	TGG	GCC	CAC	AAT	CCA	GCC	1836
Arg	Ala	Val	Leu	Lys	Leu	Met	Ser	Glu	Cys	Trp	Ala	His	Asn	Pro	Ala	
	495					500					505					

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 532 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Thr Gln Leu Tyr Ile Tyr Ile Arg Leu Leu Gly Ala Tyr Leu Phe
1 5 10 15
Ile Ile Ser Arg Val Gln Gly Gln Asn Leu Asp Ser Met Leu His Gly
20 25 30
Thr Gly Met Lys Ser Asp Ser Asp Gln Lys Lys Ser Glu Asn Gly Val
35 40 45
Thr Leu Ala Pro Glu Asp Thr Leu Pro Phe Leu Lys Cys Tyr Cys Ser
50 55 60
Gly His Cys Pro Asp Asp Ala Ile Asn Asn Thr Cys Ile Thr Asn Gly
65 70 75 80
His Cys Phe Ala Ile Ile Glu Glu Asp Asp Gln Gly Glu Thr Thr Leu
85 90 95
Ala Ser Gly Cys Met Lys Tyr Glu Gly Ser Asp Phe Gln Cys Lys Asp
100 105 110
Ser Pro Lys Ala Gln Leu Arg Arg Thr Ile Glu Cys Cys Arg Thr Asn
115 120 125
Leu Cys Asn Gln Tyr Leu Gln Pro Thr Leu Pro Pro Val Val Ile Gly
130 135 140
Pro Phe Phe Asp Gly Ser Ile Arg Trp Leu Val Leu Leu Ile Ser Met
145 150 155 160
Ala Val Cys Ile Ile Ala Met Ile Ile Phe Ser Ser Cys Phe Cys Tyr
165 170 175
Lys His Tyr Cys Lys Ser Ile Ser Ser Arg Arg Arg Tyr Asn Arg Asp
180 185 190
Leu Glu Gln Asp Glu Ala Phe Ile Pro Val Gly Glu Ser Leu Lys Asp
195 200 205
Leu Ile Asp Gln Ser Gln Ser Ser Gly Ser Gly Ser Gly Leu Pro Leu
210 215 220

SEQUENCE LISTING

Leu	Val	Gln	Arg	Thr	Ile	Ala	Lys	Gln	Ile	Gln	Met	Val	Arg	Gln	Val
225					230					235					240
Gly	Lys	Gly	Arg	Tyr	Gly	Glu	Val	Trp	Met	Gly	Lys	Trp	Arg	Gly	Glu
				245					250					255	
Lys	Val	Ala	Val	Lys	Val	Phe	Phe	Thr	Thr	Glu	Glu	Ala	Ser	Trp	Phe
			260					265					270		
Arg	Glu	Thr	Glu	Ile	Tyr	Gln	Thr	Val	Leu	Met	Arg	His	Glu	Asn	Ile
		275					280					285			
Leu	Gly	Phe	Ile	Ala	Ala	Asp	Ile	Lys	Gly	Thr	Gly	Ser	Trp	Thr	Gln
	290					295					300				
Leu	Tyr	Leu	Ile	Thr	Asp	Tyr	His	Glu	Asn	Gly	Ser	Leu	Tyr	Asp	Phe
305					310					315					320
Leu	Lys	Cys	Ala	Thr	Leu	Asp	Thr	Arg	Ala	Leu	Leu	Lys	Leu	Ala	Tyr
				325					330					335	
Ser	Ala	Ala	Cys	Gly	Leu	Cys	His	Leu	His	Thr	Glu	Ile	Tyr	Gly	Thr
			340					345					350		
Gln	Gly	Lys	Pro	Ala	Ile	Ala	His	Arg	Asp	Leu	Lys	Ser	Lys	Asn	Ile
		355					360					365			
Leu	Ile	Lys	Lys	Asn	Gly	Ser	Cys	Cys	Ile	Ala	Asp	Leu	Gly	Leu	Ala
	370					375					380				
Val	Lys	Phe	Asn	Ser	Asp	Thr	Asn	Glu	Val	Asp	Val	Pro	Leu	Asn	Thr
385					390					395					400
Arg	Val	Gly	Thr	Lys	Arg	Tyr	Met	Ala	Pro	Glu	Val	Leu	Asp	Glu	Ser
				405					410					415	
Leu	Asn	Lys	Asn	His	Phe	Gln	Pro	Tyr	Ile	Met	Ala	Asp	Ile	Tyr	Ser
			420					425					430		
Phe	Gly	Leu	Ile	Ile	Trp	Glu	Met	Ala	Arg	Arg	Cys	Ile	Thr	Gly	Gly
		435					440					445			
Ile	Val	Glu	Glu	Tyr	Gln	Leu	Pro	Tyr	Tyr	Asn	Met	Val	Pro	Ser	Asp
	450					455					460				
Pro	Ser	Tyr	Glu	Asp	Met	Arg	Glu	Val	Val	Cys	Val	Lys	Arg	Leu	Arg
465					470					475					480
Pro	Ile	Val	Ser	Asn	Arg	Trp	Asn	Ser	Asp	Glu	Cys	Leu	Arg	Ala	Val
				485					490					495	

Leu Lys Leu Met Ser Glu Cys Trp Ala His Asn Pro Ala Ser Arg Leu
 500 505 510

Thr Ala Leu Arg Ile Lys Lys Thr Leu Ala Lys Met Val Glu Ser Gln
 515 520 525

Asp Val Lys Ile
 530

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2333 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1515

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ATG GCG GAG TCG GCC GGA GCC TCC TCC TTC TTC CCC CTT GTT GTC CTC 48
 Met Ala Glu Ser Ala Gly Ala Ser Ser Phe Phe Pro Leu Val Val Leu
 1 5 10 15

CTG CTC GCC GGC AGC GGC GGG TCC GGG CCC CGG GGG GTC CAG GCT CTG 96
 Leu Leu Ala Gly Ser Gly Gly Ser Gly Pro Arg Gly Val Gln Ala Leu
 20 25 30

CTG TGT GCG TGC ACC AGC TGC CTC CAG GCC AAC TAC ACG TGT GAG ACA 144
 Leu Cys Ala Cys Thr Ser Cys Leu Gln Ala Asn Tyr Thr Cys Glu Thr
 35 40 45

GAT GGG GCC TGC ATG GTT TCC TTT TTC AAT CTG GAT GGG ATG GAG CAC 192
 Asp Gly Ala Cys Met Val Ser Phe Phe Asn Leu Asp Gly Met Glu His
 50 55 60

CAT GTG CGC ACC TGC ATC CCC AAA GTG GAG CTG GTC CCT GCC GGG AAG 240
 His Val Arg Thr Cys Ile Pro Lys Val Glu Leu Val Pro Ala Gly Lys
 65 70 75 80

"SEQUENCE" 44766060

CCC	TTC	TAC	TGC	CTG	AGC	TCG	GAG	GAC	CTG	CGC	AAC	ACC	CAC	TGC	TGC	288
Pro	Phe	Tyr	Cys	Leu	Ser	Ser	Glu	Asp	Leu	Arg	Asn	Thr	His	Cys	Cys	
			85						90					95		
TAC	ACT	GAC	TAC	TGC	AAC	AGG	ATC	GAC	TTG	AGG	GTG	CCC	AGT	GGT	CAC	336
Tyr	Thr	Asp	Tyr	Cys	Asn	Arg	Ile	Asp	Leu	Arg	Val	Pro	Ser	Gly	His	
			100					105					110			
CTC	AAG	GAG	CCT	GAG	CAC	CCG	TCC	ATG	TGG	GGC	CCG	GTG	GAG	CTG	GTA	384
Leu	Lys	Glu	Pro	Glu	His	Pro	Ser	Met	Trp	Gly	Pro	Val	Glu	Leu	Val	
		115					120					125				
GGC	ATC	ATC	GCC	GGC	CCG	GTG	TTC	CTC	CTG	TTC	CTC	ATC	ATC	ATC	ATT	432
Gly	Ile	Ile	Ala	Gly	Pro	Val	Phe	Leu	Leu	Phe	Leu	Ile	Ile	Ile	Ile	
	130					135						140				
GTT	TTC	CTT	GTC	ATT	AAC	TAT	CAT	CAG	CGT	GTC	TAT	CAC	AAC	CGC	CAG	480
Val	Phe	Leu	Val	Ile	Asn	Tyr	His	Gln	Arg	Val	Tyr	His	Asn	Arg	Gln	
					150					155					160	
AGA	CTG	GAC	ATG	GAA	GAT	CCC	TCA	TGT	GAG	ATG	TGT	CTC	TCC	AAA	GAC	528
Arg	Leu	Asp	Met	Glu	Asp	Pro	Ser	Cys	Glu	Met	Cys	Leu	Ser	Lys	Asp	
				165					170					175		
AAG	ACG	CTC	CAG	GAT	CTT	GTC	TAC	GAT	CTC	TCC	ACC	TCA	GGG	TCT	GGC	576
Lys	Thr	Leu	Gln	Asp	Leu	Val	Tyr	Asp	Leu	Ser	Thr	Ser	Gly	Ser	Gly	
			180					185					190			
TCA	GGG	TTA	CCC	CTC	TTT	GTC	CAG	CGC	ACA	GTG	GCC	CGA	ACC	ATC	GTT	624
Ser	Gly	Leu	Pro	Leu	Phe	Val	Gln	Arg	Thr	Val	Ala	Arg	Thr	Ile	Val	
		195					200					205				
TTA	CAA	GAG	ATT	ATT	GGC	AAG	GGT	CGG	TTT	GGG	GAA	GTA	TGG	CGG	GGC	672
Leu	Gln	Glu	Ile	Ile	Gly	Lys	Gly	Arg	Phe	Gly	Glu	Val	Trp	Arg	Gly	
	210					215					220					
CGC	TGG	AGG	GGT	GGT	GAT	GTG	GCT	GTG	AAA	ATA	TTC	TCT	TCT	CGT	GAA	720
Arg	Trp	Arg	Gly	Gly	Asp	Val	Ala	Val	Lys	Ile	Phe	Ser	Ser	Arg	Glu	
					230					235					240	
GAA	CGG	TCT	TGG	TTC	AGG	GAA	GCA	GAG	ATA	TAC	CAG	ACG	GTC	ATG	CTG	768
Glu	Arg	Ser	Trp	Phe	Arg	Glu	Ala	Glu	Ile	Tyr	Gln	Thr	Val	Met	Leu	
				245					250					255		
CGC	CAT	GAA	AAC	ATC	CTT	GGA	TTT	ATT	GCT	GCT	GAC	AAT	AAA	GAT	AAT	816
Arg	His	Glu	Asn	Ile	Leu	Gly	Phe	Ile	Ala	Ala	Asp	Asn	Lys	Asp	Asn	
			260					265					270			
GGC	ACC	TGG	ACA	CAG	CTG	TGG	CTT	GTT	TCT	GAC	TAT	CAT	GAG	CAC	GGG	864
Gly	Thr	Trp	Thr	Gln	Leu	Trp	Leu	Val	Ser	Asp	Tyr	His	Glu	His	Gly	
		275					280					285				

TCC Ser	CTG Leu	TTT Phe	GAT Asp	TAT Tyr	CTG Leu	AAC Asn	CGG Arg	TAC Tyr	ACA Thr	GTG Val	ACA Thr	ATT Ile	GAG Glu	GGG Gly	ATG Met	912
	290					295					300					
ATT Ile	AAG Lys	CTG Leu	GCC Ala	TTG Leu	TCT Ser	GCT Ala	GCT Ala	AGT Ser	GGG Gly	CTG Leu	GCA Ala	CAC His	CTG Leu	CAC His	ATG Met	960
305					310					315					320	
GAG Glu	ATC Ile	GTG Val	GGC Gly	ACC Thr	CAA Gln	GGG Gly	AAG Lys	CCT Pro	GGA Gly	ATT Ile	GCT Ala	CAT His	CGA Arg	GAC Asp	TTA Leu	1008
				325					330					335		
AAG Lys	TCA Ser	AAG Lys	AAC Asn	ATT Ile	CTG Leu	GTG Val	AAG Lys	AAA Lys	AAT Asn	GGC Gly	ATG Met	TGT Cys	GCC Ala	ATA Ile	GCA Ala	1056
			340					345					350			
GAC Asp	CTG Leu	GGC Gly	CTG Leu	GCT Ala	GTC Val	CGT Arg	CAT His	GAT Asp	GCA Ala	GTC Val	ACT Thr	GAC Asp	ACC Thr	ATT Ile	GAC Asp	1104
		355					360					365				
ATT Ile	GCC Ala	CCG Pro	AAT Asn	CAG Gln	AGG Arg	GTG Val	GGG Gly	ACC Thr	AAA Lys	CGA Arg	TAC Tyr	ATG Met	GCC Ala	CCT Pro	GAA Glu	1152
	370					375					380					
GTA Val	CTT Leu	GAT Asp	GAA Glu	ACC Thr	ATT Ile	AAT Asn	ATG Met	AAA Lys	CAC His	TTT Phe	GAC Asp	TCC Ser	TTT Phe	AAA Lys	TGT Cys	1200
385					390					395					400	
GCT Ala	GAT Asp	ATT Ile	TAT Tyr	GCC Ala	CTC Leu	GGG Gly	CTT Leu	GTA Val	TAT Tyr	TGG Trp	GAG Glu	ATT Ile	GCT Ala	CGA Arg	AGA Arg	1248
				405					410					415		
TGC Cys	AAT Asn	TCT Ser	GGA Gly	GGA Gly	GTC Val	CAT His	GAA Glu	GAA Glu	TAT Tyr	CAG Gln	CTG Leu	CCA Pro	TAT Tyr	TAC Tyr	GAC Asp	1296
			420				425						430			
TTA Leu	GTG Val	CCC Pro	TCT Ser	GAC Asp	CCT Pro	TCC Ser	ATT Ile	GAG Glu	GAA Glu	ATG Met	CGA Arg	AAG Lys	GTT Val	GTA Val	TGT Cys	1344
		435					440					445				
GAT Asp	CAG Gln	AAG Lys	CTG Leu	CGT Arg	CCC Pro	AAC Asn	ATC Ile	CCC Pro	AAC Asn	TGG Trp	TGG Trp	CAG Gln	AGT Ser	TAT Tyr	GAG Glu	1392
	450					455					460					
GCA Ala	CTG Leu	CGG Arg	GTG Val	ATG Met	GGG Gly	AAG Lys	ATG Met	ATG Met	CGA Arg	GAG Glu	TGT Cys	TGG Trp	TAT Tyr	GCC Ala	AAC Asn	1440
465					470					475					480	
GGC Gly	GCA Ala	GCC Ala	CGC Arg	CTG Leu	ACG Thr	GCC Ala	CTG Leu	CGC Arg	ATC Ile	AAG Lys	AAG Lys	ACC Thr	CTC Leu	TCC Ser	CAG Gln	1488
				485					490					495		

CTC AGC GTG CAG GAA GAC GTG AAG ATC TAACTGCTCC CTCTCTCCAC 1535
 Leu Ser Val Gln Glu Asp Val Lys Ile
 500 505

ACGGAGCTCC TGGCAGCGAG AACTACGCAC AGCTGCCGCG TTGAGCGTAC GATGGAGGCC 1595
 TACCTCTCGT TTCTGCCCAG CCCTCTGTGG CCAGGAGCCC TGGCCCGCAA GAGGGACAGA 1655
 GCCCGGGAGA GACTCGCTCA CTCCCATGTT GGGTTTGAGA CAGACACCTT TTCTATTAC 1715
 CTCCTAATGG CATGGAGACT CTGAGAGCGA ATTGTGTGGA GAACTCAGTG CCACACCTCG 1775
 AACTGGTTGT AGTGGGAAGT CCCGCGAAAC CCGGTGCATC TGGCACGTGG CCAGGAGCCA 1835
 TGACAGGGGC GCTTGGGAGG GGCCGGAGGA ACCGAGGTGT TGCCAGTGCT AAGCTGCCCT 1895
 GAGGGTTTCC TTCGGGGACC AGCCACAGC ACACCAAGGT GGCCCGGAAG AACCAGAAGT 1955
 GCAGCCCCTC TCACAGGCAG CTCTGAGCCG CGCTTTCCCC TCCTCCCTGG GATGGACGCT 2015
 GCCGGGAGAC TGCCAGTGGA GACGGAATCT GCCGCTTTGT CTGTCCAGCC GTGTGTGCAT 2075
 GTGCCGAGGT GCGTCCCCCG TTGTGCCTGG TTCGTGCCAT GCCCTTACAC GTGCGTGTGA 2135
 GTGTGTGTGT GTGTCTGTAG GTGCGCACTT ACCTGCTTGA GCTTTCTGTG CATGTGCAGG 2195
 TCGGGGGTGT GGTCGTCATG CTGTCCGTGC TTGCTGGTGC CTCTTTTCAG TAGTGAGCAG 2255
 CATCTAGTTT CCCTGGTGCC CTTCCCTGGA GGTCTCTCCC TCCCCCAGAG CCCCTCATGC 2315
 CACAGTGGTA CTCTGTGT 2333

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 505 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Ala Glu Ser Ala Gly Ala Ser Ser Phe Phe Pro Leu Val Val Leu
 1 5 10 15
 Leu Leu Ala Gly Ser Gly Gly Ser Gly Pro Arg Gly Val Gln Ala Leu
 20 25 30
 Leu Cys Ala Cys Thr Ser Cys Leu Gln Ala Asn Tyr Thr Cys Glu Thr
 35 40 45

Asp Gly Ala Cys Met Val Ser Phe Phe Asn Leu Asp Gly Met Glu His
 50 55 60
 His Val Arg Thr Cys Ile Pro Lys Val Glu Leu Val Pro Ala Gly Lys
 65 70 75 80
 Pro Phe Tyr Cys Leu Ser Ser Glu Asp Leu Arg Asn Thr His Cys Cys
 85 90 95
 Tyr Thr Asp Tyr Cys Asn Arg Ile Asp Leu Arg Val Pro Ser Gly His
 100 105 110
 Leu Lys Glu Pro Glu His Pro Ser Met Trp Gly Pro Val Glu Leu Val
 115 120 125
 Gly Ile Ile Ala Gly Pro Val Phe Leu Leu Phe Leu Ile Ile Ile Ile
 130 135 140
 Val Phe Leu Val Ile Asn Tyr His Gln Arg Val Tyr His Asn Arg Gln
 145 150 155 160
 Arg Leu Asp Met Glu Asp Pro Ser Cys Glu Met Cys Leu Ser Lys Asp
 165 170 175
 Lys Thr Leu Gln Asp Leu Val Tyr Asp Leu Ser Thr Ser Gly Ser Gly
 180 185 190
 Ser Gly Leu Pro Leu Phe Val Gln Arg Thr Val Ala Arg Thr Ile Val
 195 200 205
 Leu Gln Glu Ile Ile Gly Lys Gly Arg Phe Gly Glu Val Trp Arg Gly
 210 215 220
 Arg Trp Arg Gly Gly Asp Val Ala Val Lys Ile Phe Ser Ser Arg Glu
 225 230 235 240
 Glu Arg Ser Trp Phe Arg Glu Ala Glu Ile Tyr Gln Thr Val Met Leu
 245 250 255
 Arg His Glu Asn Ile Leu Gly Phe Ile Ala Ala Asp Asn Lys Asp Asn
 260 265 270
 Gly Thr Trp Thr Gln Leu Trp Leu Val Ser Asp Tyr His Glu His Gly
 275 280 285
 Ser Leu Phe Asp Tyr Leu Asn Arg Tyr Thr Val Thr Ile Glu Gly Met
 290 295 300
 Ile Lys Leu Ala Leu Ser Ala Ala Ser Gly Leu Ala His Leu His Met
 305 310 315 320

"SECRET" 44-38861-60

Glu	Ile	Val	Gly	Thr	Gln	Gly	Lys	Pro	Gly	Ile	Ala	His	Arg	Asp	Leu
			325						330					335	
Lys	Ser	Lys	Asn	Ile	Leu	Val	Lys	Lys	Asn	Gly	Met	Cys	Ala	Ile	Ala
			340					345					350		
Asp	Leu	Gly	Leu	Ala	Val	Arg	His	Asp	Ala	Val	Thr	Asp	Thr	Ile	Asp
		355					360					365			
Ile	Ala	Pro	Asn	Gln	Arg	Val	Gly	Thr	Lys	Arg	Tyr	Met	Ala	Pro	Glu
	370					375					380				
Val	Leu	Asp	Glu	Thr	Ile	Asn	Met	Lys	His	Phe	Asp	Ser	Phe	Lys	Cys
385					390					395					400
Ala	Asp	Ile	Tyr	Ala	Leu	Gly	Leu	Val	Tyr	Trp	Glu	Ile	Ala	Arg	Arg
				405					410					415	
Cys	Asn	Ser	Gly	Gly	Val	His	Glu	Glu	Tyr	Gln	Leu	Pro	Tyr	Tyr	Asp
			420					425					430		
Leu	Val	Pro	Ser	Asp	Pro	Ser	Ile	Glu	Glu	Met	Arg	Lys	Val	Val	Cys
		435					440					445			
Asp	Gln	Lys	Leu	Arg	Pro	Asn	Ile	Pro	Asn	Trp	Trp	Gln	Ser	Tyr	Glu
	450					455					460				
Ala	Leu	Arg	Val	Met	Gly	Lys	Met	Met	Arg	Glu	Cys	Trp	Tyr	Ala	Asn
465					470					475					480
Gly	Ala	Ala	Arg	Leu	Thr	Ala	Leu	Arg	Ile	Lys	Lys	Thr	Leu	Ser	Gln
				485					490					495	
Leu	Ser	Val	Gln	Glu	Asp	Val	Lys	Ile							
			500					505							

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2308 base pai

(B) TYPE: nucleic acid

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(A) ORGANISM: Mouse

(A) NAME/KEY: CDS

(B) LOCATION: 77..1585

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CGGTGGCGGC GGGACC ATG GAG GCG GCG GTC GCT GCT CCG CGT CCC CGG 109

Met Glu Ala Ala Val Ala Ala Pro Arg Pro Arg

CTG CTC CTC CTC GTG CTG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG CTG 157

CTC CCG GGG GCG ACG GCG TTA CAG TGT TTC TGC CAC CTC TGT ACA AAA 205

GAC AAT TTT ACT TGT GTG ACA GAT GGG CTC TGC TTT GTC TCT GTC ACA 253

GAG ACC ACA GAC AAA GTT ATA CAC AAC AGC ATG TGT ATA GCT GAA ATT 301

GAC TTA ATT CCT CGA GAT AGG CCG TTT GTA TGT GCA CCC TCT TCA AAA 349

ACT GGG TCT GTG ACT ACA ACA TAT TGC TGC AAT CAG GAC CAT TGC AAT 397

AAA ATA GAA CTT CCA ACT ACT GTA AAG TCA TCA CCT GGC CTT GGT CCT 445

GTG Val	GAA Glu	CTG Leu	GCA Ala	GCT Ala	GTC Val	ATT Ile	GCT Ala	GGA Gly	CCA Pro	GTG Val	TGC Cys	TTC Phe	GTC Val	TGC Cys	ATC Ile	493
	125					130					135					
TCA Ser	CTC Leu	ATG Met	TTG Leu	ATG Met	GTC Val	TAT Tyr	ATC Ile	TGC Cys	CAC His	AAC Asn	CGC Arg	ACT Thr	GTC Val	ATT Ile	CAC His	541
	140				145					150					155	
CAT His	CGA Arg	GTG Val	CCA Pro	AAT Asn	GAA Glu	GAG Glu	GAC Asp	CCT Pro	TCA Ser	TTA Leu	GAT Asp	CGC Arg	CCT Pro	TTT Phe	ATT Ile	589
				160					165					170		
TCA Ser	GAG Glu	GGT Gly	ACT Thr	ACG Thr	TTG Leu	AAA Lys	GAC Asp	TTA Leu	ATT Ile	TAT Tyr	GAT Asp	ATG Met	ACA Thr	ACG Thr	TCA Ser	637
			175					180					185			
GGT Gly	TCT Ser	GGC Gly	TCA Ser	GGT Gly	TTA Leu	CCA Pro	TTG Leu	CTT Leu	GTT Val	CAG Gln	AGA Arg	ACA Thr	ATT Ile	GCG Ala	AGA Arg	685
		190					195					200				
ACT Thr	ATT Ile	GTG Val	TTA Leu	CAA Gln	GAA Glu	AGC Ser	ATT Ile	GGC Gly	AAA Lys	GGT Gly	CGA Arg	TTT Phe	GGA Gly	GAA Glu	GTT Val	733
	205					210					215					
TGG Trp	AGA Arg	GGA Gly	AAG Lys	TGG Trp	CGG Arg	GGA Gly	GAA Glu	GAA Glu	GTT Val	GCT Ala	GTT Val	AAG Lys	ATA Ile	TTC Phe	TCC Ser	781
	220				225					230					235	
TCT Ser	AGA Arg	GAA Glu	GAA Glu	CGT Arg	TCG Ser	TGG Trp	TTC Phe	CGT Arg	GAG Glu	GCA Ala	GAG Glu	ATT Ile	TAT Tyr	CAA Gln	ACT Thr	829
				240					245					250		
GTA Val	ATG Met	TTA Leu	CGT Arg	CAT His	GAA Glu	AAC Asn	ATC Ile	CTG Leu	GGA Gly	TTT Phe	ATA Ile	GCA Ala	GCA Ala	GAC Asp	AAT Asn	877
			255					260				265				
AAA Lys	GAC Asp	AAT Asn	GGT Gly	ACT Thr	TGG Trp	ACT Thr	CAG Gln	CTC Leu	TGG Trp	TTG Leu	GTG Val	TCA Ser	GAT Asp	TAT Tyr	CAT His	925
		270					275					280				
GAG Glu	CAT His	GGA Gly	TCC Ser	CTT Leu	TTT Phe	GAT Asp	TAC Tyr	TTA Leu	AAC Asn	AGA Arg	TAC Tyr	ACA Thr	GTT Val	ACT Thr	GTG Val	973
	285					290					295					
GAA Glu	GGA Gly	ATG Met	ATA Ile	AAA Lys	CTT Leu	GCT Ala	CTG Leu	TCC Ser	ACG Thr	GCG Ala	AGC Ser	GGT Gly	CTT Leu	GCC Ala	CAT His	1021
	300				305					310					315	
CTT Leu	CAC His	ATG Met	GAG Glu	ATT Ile	GTT Val	GGT Gly	ACC Thr	CAA Gln	GGA Gly	AAG Lys	CCA Pro	GCC Ala	ATT Ile	GCT Ala	CAT His	1069
				320					325					330		

[illegible]

(2) INFORMATION FOR SEQ ID NO: 10:

(A) LENGTH: 503 amino acids

(D) TOPOLOGY: linear

MOLECULE TYPE: prote

(xi) SEQUENCE DESCRIPTION:

(11) SEQUENCE NUMBER: SEQ ID NO: 10:

Met	Glu	Ala	Ala	Val	Ala	Ala	Pro	Arg	Pro	Arg	Leu	Leu	Leu	Leu	Val
1				5					10					15	
Leu	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Leu	Leu	Pro	Gly	Ala	Thr
			20					25					30		
Ala	Leu	Gln	Cys	Phe	Cys	His	Leu	Cys	Thr	Lys	Asp	Asn	Phe	Thr	Cys
		35					40					45			
Val	Thr	Asp	Gly	Leu	Cys	Phe	Val	Ser	Val	Thr	Glu	Thr	Thr	Asp	Lys
	50					55					60				
Val	Ile	His	Asn	Ser	Met	Cys	Ile	Ala	Glu	Ile	Asp	Leu	Ile	Pro	Arg
65					70					75					80
Asp	Arg	Pro	Phe	Val	Cys	Ala	Pro	Ser	Ser	Lys	Thr	Gly	Ser	Val	Thr
				85					90					95	
Thr	Thr	Tyr	Cys	Cys	Asn	Gln	Asp	His	Cys	Asn	Lys	Ile	Glu	Leu	Pro
			100					105					110		
Thr	Thr	Val	Lys	Ser	Ser	Pro	Gly	Leu	Gly	Pro	Val	Glu	Leu	Ala	Ala
		115					120					125			

Val 130	Ile	Ala	Gly	Pro	Val 135	Cys	Phe	Val	Cys	Ile	Ser 140	Leu	Met	Leu	Met
Val 145	Tyr	Ile	Cys	His	Asn 150	Arg	Thr	Val	Ile	His 155	His	Arg	Val	Pro	Asn 160
Glu	Glu	Asp	Pro	Ser 165	Leu	Asp	Arg	Pro	Phe 170	Ile	Ser	Glu	Gly	Thr 175	Thr
Leu	Lys	Asp	Leu 180	Ile	Tyr	Asp	Met	Thr 185	Thr	Ser	Gly	Ser	Gly 190	Ser	Gly
Leu	Pro	Leu 195	Leu	Val	Gln	Arg	Thr 200	Ile	Ala	Arg	Thr	Ile 205	Val	Leu	Gln
Glu	Ser 210	Ile	Gly	Lys	Gly	Arg 215	Phe	Gly	Glu	Val	Trp 220	Arg	Gly	Lys	Trp
Arg 225	Gly	Glu	Glu	Val	Ala 230	Val	Lys	Ile	Phe	Ser 235	Ser	Arg	Glu	Glu	Arg 240
Ser	Trp	Phe	Arg	Glu 245	Ala	Glu	Ile	Tyr	Gln 250	Thr	Val	Met	Leu	Arg 255	His
Glu	Asn	Ile	Leu 260	Gly	Phe	Ile	Ala	Ala 265	Asp	Asn	Lys	Asp	Asn 270	Gly	Thr
Trp	Thr	Gln 275	Leu	Trp	Leu	Val	Ser 280	Asp	Tyr	His	Glu	His 285	Gly	Ser	Leu
Phe	Asp 290	Tyr	Leu	Asn	Arg	Tyr 295	Thr	Val	Thr	Val	Glu 300	Gly	Met	Ile	Lys
Leu 305	Ala	Leu	Ser	Thr	Ala 310	Ser	Gly	Leu	Ala	His 315	Leu	His	Met	Glu	Ile 320
Val	Gly	Thr	Gln	Gly 325	Lys	Pro	Ala	Ile	Ala 330	His	Arg	Asp	Leu	Lys 335	Ser
Lys	Asn	Ile	Leu 340	Val	Lys	Lys	Asn	Gly 345	Thr	Cys	Cys	Ile	Ala 350	Asp	Leu
Gly	Leu	Ala 355	Val	Arg	His	Asp	Ser 360	Ala	Thr	Asp	Thr	Ile 365	Asp	Ile	Ala
Pro	Asn 370	His	Arg	Val	Gly	Thr 375	Lys	Arg	Tyr	Met	Ala 380	Pro	Glu	Val	Leu
Asp 385	Asp	Ser	Ile	Asn	Met 390	Lys	His	Phe	Glu	Ser 395	Phe	Lys	Arg	Ala	Asp 400

Ile Tyr Ala Met Gly Leu Val Phe Trp Glu Ile Ala Arg Arg Cys Ser
 405 410 415

Ile Gly Gly Ile His Glu Asp Tyr Gln Leu Pro Tyr Tyr Asp Leu Val
 420 425 430

Pro Ser Asp Pro Ser Val Glu Glu Met Arg Lys Val Val Cys Glu Gln
 435 440 445

Lys Leu Arg Pro Asn Ile Pro Asn Arg Trp Gln Ser Cys Glu Ala Leu
 450 455 460

Arg Val Met Ala Lys Ile Met Arg Glu Cys Trp Tyr Ala Asn Gly Ala
 465 470 475 480

Ala Arg Leu Thr Ala Leu Arg Ile Lys Lys Thr Leu Ser Gln Leu Ser
 485 490 495

Gln Gln Glu Gly Ile Lys Met
 500

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1922 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mouse

(ix) FEATURE:

- (A) NAME/KEY: CDS

- (B) LOCATION: 241..1746

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GAGAGCACAG CCCTTCCCAG TCCCCGGAGC CGCCGCGCCA CGCGCGCATG ATCAAGACCT	60
TTTCCCCGGC CCCACAGGGC CTCTGGACGT GAGACCCCGG CCGCCTCCGC AAGGAGAGGC	120
GGGGGTCGAG TCGCCCTGTC CAAAGGCCTC AATCTAAACA ATCTTGATTC CTGTTGCCGG	180
CTGGCGGGAC CCTGAATGGC AGGAAATCTC ACCACATCTC TTCTCCTATC TCCAAGGACC	240
ATG ACC TTG GGG AGC TTC AGA AGG GGC CTT TTG ATG CTG TCG GTG GCC	288
Met Thr Leu Gly Ser Phe Arg Arg Gly Leu Leu Met Leu Ser Val Ala	
1 5 10 15	

"SEQUENCE" 06/06/00

GTG Val 225	GCG Ala	GTC Val	AAG Lys	ATT Ile	TTC Phe 230	TCC Ser	TCA Ser	CGA Arg	GAT Asp	GAG Glu 235	CAG Gln	TCC Ser	TGG Trp	TTC Phe	CGG Arg 240	960
GAG Glu	ACG Thr	GAG Glu	ATC Ile	TAC Tyr 245	AAC Asn	ACA Thr	GTT Val	CTG Leu	CTT Leu 250	AGA Arg	CAC His	GAC Asp	AAC Asn	ATC Ile 255	CTA Leu	1008
GGC Gly	TTC Phe	ATC Ile	GCC Ala 260	TCC Ser	GAC Asp	ATG Met	ACT Thr	TCG Ser 265	CGG Arg	AAC Asn	TCG Ser	AGC Ser	ACG Thr 270	CAG Gln	CTG Leu	1056
TGG Trp	CTC Leu	ATC Ile 275	ACC Thr	CAC His	TAC Tyr	CAT His	GAA Glu 280	CAC His	GGC Gly	TCC Ser	CTC Leu	TAT Tyr 285	GAC Asp	TTT Phe	CTG Leu	1104
CAG Gln 290	AGG Arg	CAG Gln	ACG Thr	CTG Leu	GAG Glu	CCC Pro 295	CAG Gln	TTG Leu	GCC Ala	CTG Leu	AGG Arg 300	CTA Leu	GCT Ala	GTG Val	TCC Ser	1152
CCG Pro 305	GCC Ala	TGC Cys	GGC Gly	CTG Leu	GCG Ala 310	CAC His	CTA Leu	CAT His	GTG Val	GAG Glu 315	ATC Ile	TTT Phe	GGC Gly	ACT Thr	CAA Gln 320	1200
GGC Gly	AAA Lys	CCA Pro	GCC Ala	ATT Ile 325	GCC Ala	CAT His	CGT Arg	GAC Asp	CTC Leu 330	AAG Lys	AGT Ser	CGC Arg	AAT Asn	GTG Val 335	CTG Leu	1248
GTC Val	AAG Lys	AGT Ser	AAC Asn 340	TTG Leu	CAG Gln	TGT Cys	TGC Cys	ATT Ile 345	GCA Ala	GAC Asp	CTG Leu	GGA Gly	CTG Leu 350	GCT Ala	GTG Val	1296
ATG Met	CAC His	TCA Ser 355	CAA Gln	AGC Ser	AAC Asn	GAG Glu	TAC Tyr 360	CTG Leu	GAT Asp	ATC Ile	GGC Gly	AAC Asn 365	ACA Thr	CCC Pro	CGA Arg	1344
GTG Val 370	GGT Gly	ACC Thr	AAA Lys	AGA Arg	TAC Tyr	ATG Met 375	GCA Ala	CCC Pro	GAG Glu	GTG Val	CTG Leu 380	GAT Asp	GAG Glu	CAC His	ATC Ile	1392
CGC Arg 385	ACA Thr	GAC Asp	TGC Cys	TTT Phe	GAG Glu 390	TCG Ser	TAC Tyr	AAG Lys	TGG Trp	ACA Thr 395	GAC Asp	ATC Ile	TGG Trp	GCC Ala	TTT Phe 400	1440
GGC Gly	CTA Leu	GTG Val	CTA Leu	TGG Trp 405	GAG Glu	ATC Ile	GCC Ala	CGG Arg	CGG Arg 410	ACC Thr	ATC Ile	ATC Ile	AAT Asn	GGC Gly 415	ATT Ile	1488
GTG Val	GAG Glu	GAT Asp	TAC Tyr 420	AGG Arg	CCA Pro	CCT Pro	TTC Phe	TAT Tyr 425	GAC Asp	ATG Met	GTA Val	CCC Pro	AAT Asn 430	GAC Asp	CCC Pro	1536

AGT TTT GAG GAC ATG AAA AAG GTG GTG TGC GTT GAC CAG CAG ACA CCC 1584
 Ser Phe Glu Asp Met Lys Lys Val Val Cys Val Asp Gln Gln Thr Pro
 435 440 445

ACC ATC CCT AAC CGG CTG GCT GCA GAT CCG GTC CTC TCC GGG CTG GCC 1632
 Thr Ile Pro Asn Arg Leu Ala Ala Asp Pro Val Leu Ser Gly Leu Ala
 450 455 460

CAG ATG ATG AGA GAG TGC TGG TAC CCC AAC CCC TCT GCT CGC CTC ACC 1680
 Gln Met Met Arg Glu Cys Trp Tyr Pro Asn Pro Ser Ala Arg Leu Thr
 465 470 475 480

GCA CTG CGC ATA AAG AAG ACA TTG CAG AAG CTC AGT CAC AAT CCA GAG 1728
 Ala Leu Arg Ile Lys Lys Thr Leu Gln Lys Leu Ser His Asn Pro Glu
 485 490 495

AAG CCC AAA GTG ATT CAC TAGCCCAGGG CCACCAGGCT TCCTCTGCCT 1776
 Lys Pro Lys Val Ile His
 500

AAAGTGTGTG CTGGGGAAGA AGACATAGCC TGTCTGGGTA GAGGGAGTGA AGAGAGTGTG 1836
 CACGCTGCCC TGTGTGTGCC TGCTCAGCTT GCTCCCAGCC CATCCAGCCA AAAATACAGC 1896
 TGAGCTGAAA TTCAAAAAAA AAAAAA 1922

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 502 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Thr Leu Gly Ser Phe Arg Arg Gly Leu Leu Met Leu Ser Val Ala
 1 5 10 15

Leu Gly Leu Thr Gln Gly Arg Leu Ala Lys Pro Ser Lys Leu Val Asn
 20 25 30

Cys Thr Cys Glu Ser Pro His Cys Lys Arg Pro Phe Cys Gln Gly Ser
 35 40 45

Trp Cys Thr Val Val Leu Val Arg Glu Gln Gly Arg His Pro Gln Val
 50 55 60

Tyr Arg Gly Cys Gly Ser Leu Asn Gln Glu Leu Cys Leu Gly Arg Pro
 65 70 75 80

SECRETED

Thr	Glu	Phe	Leu	Asn	His	His	Cys	Cys	Tyr	Arg	Ser	Phe	Cys	Asn	His
				85					90					95	
Asn	Val	Ser	Leu	Met	Leu	Glu	Ala	Thr	Gln	Thr	Pro	Ser	Glu	Glu	Pro
			100					105					110		
Glu	Val	Asp	Ala	His	Leu	Pro	Leu	Ile	Leu	Gly	Pro	Val	Leu	Ala	Leu
		115					120					125			
Pro	Val	Leu	Val	Ala	Leu	Gly	Ala	Leu	Gly	Leu	Trp	Arg	Val	Arg	Arg
	130					135					140				
Arg	Gln	Glu	Lys	Gln	Arg	Asp	Leu	His	Ser	Asp	Leu	Gly	Glu	Ser	Ser
145					150					155					160
Leu	Ile	Leu	Lys	Ala	Ser	Glu	Gln	Ala	Asp	Ser	Met	Leu	Gly	Asp	Phe
				165					170					175	
Leu	Asp	Ser	Asp	Cys	Thr	Thr	Gly	Ser	Gly	Ser	Gly	Leu	Pro	Phe	Leu
			180					185					190		
Val	Gln	Arg	Thr	Val	Ala	Arg	Gln	Val	Ala	Leu	Val	Glu	Cys	Val	Gly
		195					200					205			
Lys	Gly	Arg	Tyr	Gly	Glu	Val	Trp	Arg	Gly	Ser	Trp	His	Gly	Glu	Ser
	210					215					220				
Val	Ala	Val	Lys	Ile	Phe	Ser	Ser	Arg	Asp	Glu	Gln	Ser	Trp	Phe	Arg
225					230					235					240
Glu	Thr	Glu	Ile	Tyr	Asn	Thr	Val	Leu	Leu	Arg	His	Asp	Asn	Ile	Leu
				245					250					255	
Gly	Phe	Ile	Ala	Ser	Asp	Met	Thr	Ser	Arg	Asn	Ser	Ser	Thr	Gln	Leu
			260					265					270		
Trp	Leu	Ile	Thr	His	Tyr	His	Glu	His	Gly	Ser	Leu	Tyr	Asp	Phe	Leu
		275					280					285			
Gln	Arg	Gln	Thr	Leu	Glu	Pro	Gln	Leu	Ala	Leu	Arg	Leu	Ala	Val	Ser
	290					295					300				
Pro	Ala	Cys	Gly	Leu	Ala	His	Leu	His	Val	Glu	Ile	Phe	Gly	Thr	Gln
305					310					315					320
Gly	Lys	Pro	Ala	Ile	Ala	His	Arg	Asp	Leu	Lys	Ser	Arg	Asn	Val	Leu
				325					330					335	
Val	Lys	Ser	Asn	Leu	Gln	Cys	Cys	Ile	Ala	Asp	Leu	Gly	Leu	Ala	Val
			340					345					350		

Met His Ser Gln Ser Asn Glu Tyr Leu Asp Ile Gly Asn Thr Pro Arg
 355 360 365
 Val Gly Thr Lys Arg Tyr Met Ala Pro Glu Val Leu Asp Glu His Ile
 370 375 380
 Arg Thr Asp Cys Phe Glu Ser Tyr Lys Trp Thr Asp Ile Trp Ala Phe
 385 390 395 400
 Gly Leu Val Leu Trp Glu Ile Ala Arg Arg Thr Ile Ile Asn Gly Ile
 405 410 415
 Val Glu Asp Tyr Arg Pro Pro Phe Tyr Asp Met Val Pro Asn Asp Pro
 420 425 430
 Ser Phe Glu Asp Met Lys Lys Val Val Cys Val Asp Gln Gln Thr Pro
 435 440 445
 Thr Ile Pro Asn Arg Leu Ala Ala Asp Pro Val Leu Ser Gly Leu Ala
 450 455 460
 Gln Met Met Arg Glu Cys Trp Tyr Pro Asn Pro Ser Ala Arg Leu Thr
 465 470 475 480
 Ala Leu Arg Ile Lys Lys Thr Leu Gln Lys Leu Ser His Asn Pro Glu
 485 490 495
 Lys Pro Lys Val Ile His
 500

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2070 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mouse

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 217..1812

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATTCATGAGA TGGAAGCATA GGTCAAAGCT GTTCGGAGAA ATTGGA ACTA CAGTTTTATC

TAGCCACATC	TCTGAGAATT	CTGAAGAAAG	CAGCAGGTGA	AAGTCATTGC	CAAGTGATTT	120
TGTTCTGTAA	GGAAGCCTCC	CTCATTCACT	TACACCAGTG	AGACAGCAGG	ACCAGTCATT	180
CAAAGGGCCG	TGTACAGGAC	GCGTGGCAAT	CAGACA	ATG	ACT	234
				Met	Thr	
				1		
TAC	ATC	AGA	TTA	CTG	GGA	282
Tyr	Ile	Arg	Leu	Leu	Gly	
			10			
GGG	CAG	AAT	CTA	GAT	AGT	330
Gly	Gln	Asn	Leu	Asp	Ser	
		25				
TTG	GAC	CAG	AAG	AAG	CCA	378
Leu	Asp	Gln	Lys	Lys	Pro	
	40					
ACC	TTG	CCT	TTC	TTA	AAG	426
Thr	Leu	Pro	Phe	Leu	Lys	
55					60	
GCT	ATT	AAT	AAC	ACA	TGC	474
Ala	Ile	Asn	Asn	Thr	Cys	
				75		
GAA	GAA	GAT	GAT	CAG	GGA	522
Glu	Glu	Asp	Asp	Gln	Gly	
			90			
TAT	GAA	GGC	TCT	GAT	TTT	570
Tyr	Glu	Gly	Ser	Asp	Phe	
		105				
CGC	AGG	ACA	ATA	GAA	TGT	618
Arg	Arg	Thr	Ile	Glu	Cys	
	120					
CAG	CCT	ACA	CTG	CCC	CCT	666
Gln	Pro	Thr	Leu	Pro	Pro	
135					140	
ATC	CGA	TGG	CTG	GTT	GTG	714
Ile	Arg	Trp	Leu	Val	Val	
				155		
ATG	ATC	ATC	TTC	TCC	AGC	762
Met	Ile	Ile	Phe	Ser	Ser	
			170			

ATC Ile	TCA Ser	AGC Ser 185	AGG Arg	GGT Gly	CGT Arg	TAC Tyr	AAC Asn 190	CGT Arg	GAT Asp	TTG Leu	GAA Glu	CAG Gln 195	GAT Asp	GAA Glu	GCA Ala	810
TTT Phe	ATT Ile 200	CCA Pro	GTA Val	GGA Gly	GAA Glu	TCA Ser 205	TTG Leu	AAA Lys	GAC Asp	CTG Leu	ATT Ile 210	GAC Asp	CAG Gln	TCC Ser	CAA Gln	858
AGC Ser 215	TCT Ser	GGG Gly	AGT Ser	GGA Gly	TCT Ser 220	GGA Gly	TTG Leu	CCT Pro	TTA Leu	TTG Leu 225	GTT Val	CAG Gln	CGA Arg	ACT Thr	ATT Ile 230	906
GCC Ala	AAA Lys	CAG Gln	ATT Ile	CAG Gln 235	ATG Met	GTT Val	CGG Arg	CAG Gln	GTT Val 240	GGT Gly	AAA Lys	GGC Gly	CGC Arg	TAT Tyr 245	GGA Gly	954
GAA Glu	GTA Val	TGG Trp 250	ATG Met	GGT Gly	AAA Lys	TGG Trp	CGT Arg	GGT Gly 255	GAA Glu	AAA Lys	GTG Val	GCT Ala	GTC Val 260	AAA Lys	GTG Val	1002
TTT Phe	TTT Phe	ACC Thr 265	ACT Thr	GAA Glu	GAA Glu	GCT Ala	AGC Ser 270	TGG Trp	TTT Phe	AGA Arg	GAA Glu	ACA Thr 275	GAA Glu	ATC Ile	TAC Tyr	1050
CAG Gln	ACG Thr 280	GTG Val	TTA Leu	ATG Met	CGT Arg	CAT His 285	GAA Glu	AAT Asn	ATA Ile	CTT Leu	GGT Gly 290	TTT Phe	ATA Ile	GCT Ala	GCA Ala	1098
GAC Asp 295	ATT Ile	AAA Lys	GGC Gly	ACT Thr	GGT Gly 300	TCC Ser	TGG Trp	ACT Thr	CAG Gln	CTG Leu 305	TAT Tyr	TTG Leu	ATT Ile	ACT Thr	GAT Asp 310	1146
TAC Tyr	CAT His	GAA Glu	AAT Asn	GGA Gly 315	TCT Ser	CTC Leu	TAT Tyr	GAC Asp	TTC Phe 320	CTG Leu	AAA Lys	TGT Cys	GCC Ala	ACA Thr 325	CTA Leu	1194
GAC Asp	ACC Thr	AGA Arg	GCC Ala 330	CTA Leu	CTC Leu	AAG Lys	TTA Leu	GCT Ala 335	TAT Tyr	TCT Ser	GCT Ala	GCT Ala	TGT Cys 340	GGT Gly	CTG Leu	1242
TGC Cys	CAC His	CTC Leu 345	CAC His	ACA Thr	GAA Glu	ATT Ile	TAT Tyr 350	GGT Gly	ACC Thr	CAA Gln	GGG Gly	AAG Lys 355	CCT Pro	GCA Ala	ATT Ile	1290
GCT Ala	CAT His 360	CGA Arg	GAC Asp	CTG Leu	AAG Lys	AGC Ser 365	AAA Lys	AAC Asn	ATC Ile	CTT Leu	ATT Ile 370	AAG Lys	AAA Lys	AAT Asn	GGA Gly	1338
AGT Ser 375	TGC Cys	TGT Cys	ATT Ile	GCT Ala	GAC Asp 380	CTG Leu	GGC Gly	CTA Leu	GCT Ala	GTT Val 385	AAA Lys	TTC Phe	AAC Asn	AGT Ser	GAT Asp 390	1386

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Leu Ile Asp Gln Ser Gln Ser Ser Gly Ser Gly Ser Gly Leu Pro Leu
210 215 220

Leu 225	Val	Gln	Arg	Thr	Ile 230	Ala	Lys	Gln	Ile	Gln 235	Met	Val	Arg	Gln	Val 240
Gly	Lys	Gly	Arg	Tyr 245	Gly	Glu	Val	Trp	Met 250	Gly	Lys	Trp	Arg	Gly 255	Glu
Lys	Val	Ala	Val 260	Lys	Val	Phe	Phe	Thr 265	Thr	Glu	Glu	Ala	Ser 270	Trp	Phe
Arg	Glu	Thr 275	Glu	Ile	Tyr	Gln	Thr 280	Val	Leu	Met	Arg	His 285	Glu	Asn	Ile
Leu 290	Gly	Phe	Ile	Ala	Ala	Asp 295	Ile	Lys	Gly	Thr	Gly 300	Ser	Trp	Thr	Gln
Leu 305	Tyr	Leu	Ile	Thr	Asp 310	Tyr	His	Glu	Asn	Gly 315	Ser	Leu	Tyr	Asp	Phe 320
Leu	Lys	Cys	Ala	Thr 325	Leu	Asp	Thr	Arg	Ala 330	Leu	Leu	Lys	Leu	Ala 335	Tyr
Ser	Ala	Ala	Cys 340	Gly	Leu	Cys	His	Leu 345	His	Thr	Glu	Ile	Tyr 350	Gly	Thr
Gln	Gly	Lys 355	Pro	Ala	Ile	Ala	His 360	Arg	Asp	Leu	Lys	Ser 365	Lys	Asn	Ile
Leu 370	Ile	Lys	Lys	Asn	Gly	Ser 375	Cys	Cys	Ile	Ala	Asp 380	Leu	Gly	Leu	Ala
Val 385	Lys	Phe	Asn	Ser	Asp 390	Thr	Asn	Glu	Val	Asp 395	Ile	Pro	Leu	Asn	Thr 400
Arg	Val	Gly	Thr 405	Lys	Arg	Tyr	Met	Ala	Pro 410	Glu	Val	Leu	Asp	Glu 415	Ser
Leu	Asn	Lys	Asn 420	His	Phe	Gln	Pro	Tyr 425	Ile	Met	Ala	Asp	Ile 430	Tyr	Ser
Phe	Gly	Leu 435	Ile	Ile	Trp	Glu	Met 440	Ala	Arg	Arg	Cys	Ile 445	Thr	Gly	Gly
Ile 450	Val	Glu	Glu	Tyr	Gln	Leu 455	Pro	Tyr	Tyr	Asn	Met 460	Val	Pro	Ser	Asp
Pro 465	Ser	Tyr	Glu	Asp	Met 470	Arg	Glu	Val	Val	Cys 475	Val	Lys	Arg	Leu	Arg 480
Pro	Ile	Val	Ser	Asn 485	Arg	Trp	Asn	Ser	Asp 490	Glu	Cys	Leu	Arg	Ala 495	Val

Leu Lys Leu Met Ser Glu Cys Trp Ala His Asn Pro Ala Ser Arg Leu
500 505 510

Thr Ala Leu Arg Ile Lys Lys Thr Leu Ala Lys Met Val Glu Ser Gln
515 520 525

Asp Val Lys Ile
530

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2160 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mouse

(ix) FEATURE:

- (A) NAME/KEY: CDS

- (B) LOCATION: 10..1524

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CGCGGTTAC ATG GCG GAG TCG GCC GGA GCC TCC TCC TTC TTC CCC CTT 48
Met Ala Glu Ser Ala Gly Ala Ser Ser Phe Phe Pro Leu
1 5 10

GTT GTC CTC CTG CTC GCC GGC AGC GGC GGG TCC GGG CCC CGG GGG ATC 96
Val Val Leu Leu Leu Ala Gly Ser Gly Gly Ser Gly Pro Arg Gly Ile
15 20 25

CAG GCT CTG CTG TGT GCG TGC ACC AGC TGC CTA CAG ACC AAC TAC ACC 144
Gln Ala Leu Leu Cys Ala Cys Thr Ser Cys Leu Gln Thr Asn Tyr Thr
30 35 40 45

TGT GAG ACA GAT GGG GCT TGC ATG GTC TCC ATC TTT AAC CTG GAT GGC 192
Cys Glu Thr Asp Gly Ala Cys Met Val Ser Ile Phe Asn Leu Asp Gly
50 55 60

GTG GAG CAC CAT GTA CGT ACC TGC ATC CCC AAG GTG GAG CTG GTT CCT 240
Val Glu His His Val Arg Thr Cys Ile Pro Lys Val Glu Leu Val Pro
65 70 75

GCT GGA AAG CCC TTC TAC TGC CTG AGT TCA GAG GAT CTG CGC AAC ACA 288
Ala Gly Lys Pro Phe Tyr Cys Leu Ser Ser Glu Asp Leu Arg Asn Thr
80 85 90

SEQUENCE 15

CAC	TGC	TGC	TAT	ATT	GAC	TTC	TGC	AAC	AAG	ATT	GAC	CTC	AGG	GTC	CCC	336
His	Cys	Cys	Tyr	Ile	Asp	Phe	Cys	Asn	Lys	Ile	Asp	Leu	Arg	Val	Pro	
	95					100					105					
AGC	GGA	CAC	CTC	AAG	GAG	CCT	GCG	CAC	CCC	TCC	ATG	TGG	GGC	CCT	GTG	384
Ser	Gly	His	Leu	Lys	Glu	Pro	Ala	His	Pro	Ser	Met	Trp	Gly	Pro	Val	
110					115					120					125	
GAG	CTG	GTC	GGC	ATC	ATC	GCC	GGC	CCC	GTC	TTC	CTC	CTC	TTC	CTT	ATC	432
Glu	Leu	Val	Gly	Ile	Ile	Ala	Gly	Pro	Val	Phe	Leu	Leu	Phe	Leu	Ile	
				130					135					140		
ATT	ATC	ATC	GTC	TTC	CTG	GTC	ATC	AAC	TAT	CAC	CAG	CGT	GTC	TAC	CAT	480
Ile	Ile	Ile	Val	Phe	Leu	Val	Ile	Asn	Tyr	His	Gln	Arg	Val	Tyr	His	
			145					150					155			
AAC	CGC	CAG	AGG	TTG	GAC	ATG	GAG	GAC	CCC	TCT	TGC	GAG	ATG	TGT	CTC	528
Asn	Arg	Gln	Arg	Leu	Asp	Met	Glu	Asp	Pro	Ser	Cys	Glu	Met	Cys	Leu	
		160					165					170				
TCC	AAA	GAC	AAG	ACG	CTC	CAG	GAT	CTC	GTC	TAC	GAC	CTC	TCC	ACG	TCA	576
Ser	Lys	Asp	Lys	Thr	Leu	Gln	Asp	Leu	Val	Tyr	Asp	Leu	Ser	Thr	Ser	
	175					180					185					
GGG	TCT	GGC	TCA	GGG	TTA	CCC	CTT	TTT	GTC	CAG	CGC	ACA	GTG	GCC	CGA	624
Gly	Ser	Gly	Ser	Gly	Leu	Pro	Leu	Phe	Val	Gln	Arg	Thr	Val	Ala	Arg	
190					195					200					205	
ACC	ATT	GTT	TTA	CAA	GAG	ATT	ATC	GGC	AAG	GGC	CGG	TTC	GGG	GAA	GTA	672
Thr	Ile	Val	Leu	Gln	Glu	Ile	Ile	Gly	Lys	Gly	Arg	Phe	Gly	Glu	Val	
				210					215					220		
TGG	CGT	GGT	CGC	TGG	AGG	GGT	GGT	GAC	GTG	GCT	GTG	AAA	ATC	TTC	TCT	720
Trp	Arg	Gly	Arg	Trp	Arg	Gly	Gly	Asp	Val	Ala	Val	Lys	Ile	Phe	Ser	
			225					230					235			
TCT	CGT	GAA	GAA	CGG	TCT	TGG	TTC	CGT	GAA	GCA	GAG	ATC	TAC	CAG	ACC	768
Ser	Arg	Glu	Glu	Arg	Ser	Trp	Phe	Arg	Glu	Ala	Glu	Ile	Tyr	Gln	Thr	
		240					245					250				
GTC	ATG	CTG	CGC	CAT	GAA	AAC	ATC	CTT	GGC	TTT	ATT	GCT	GCT	GAC	AAT	816
Val	Met	Leu	Arg	His	Glu	Asn	Ile	Leu	Gly	Phe	Ile	Ala	Ala	Asp	Asn	
	255					260					265					
AAA	GAT	AAT	GGC	ACC	TGG	ACC	CAG	CTG	TGG	CTT	GTC	TCT	GAC	TAT	CAC	864
Lys	Asp	Asn	Gly	Thr	Trp	Thr	Gln	Leu	Trp	Leu	Val	Ser	Asp	Tyr	His	
270					275					280					285	

090391 091398

GAG	CAT	GGC	TCA	CTG	TTT	GAT	TAT	CTG	AAC	CGC	TAC	ACA	GTG	ACC	ATT	912
Glu	His	Gly	Ser	Leu	Phe	Asp	Tyr	Leu	Asn	Arg	Tyr	Thr	Val	Thr	Ile	
				290					295					300		
GAG	GGA	ATG	ATT	AAG	CTA	GCC	TTG	TCT	GCA	GCC	AGT	GGT	TTG	GCA	CAC	960
Glu	Gly	Met	Ile	Lys	Leu	Ala	Leu	Ser	Ala	Ala	Ser	Gly	Leu	Ala	His	
			305					310					315			
CTG	CAT	ATG	GAG	ATT	GTG	GGC	ACT	CAA	GGG	AAG	CCG	GGA	ATT	GCT	CAT	1008
Leu	His	Met	Glu	Ile	Val	Gly	Thr	Gln	Gly	Lys	Pro	Gly	Ile	Ala	His	
		320					325					330				
CGA	GAC	TTG	AAG	TCA	AAG	AAC	ATC	CTG	GTG	AAA	AAA	AAT	GGC	ATG	TGT	1056
Arg	Asp	Leu	Lys	Ser	Lys	Asn	Ile	Leu	Val	Lys	Lys	Asn	Gly	Met	Cys	
	335					340					345					
GCC	ATT	GCA	GAC	CTG	GGC	CTG	GCT	GTC	CGT	CAT	GAT	GCG	GTC	ACT	GAC	1104
Ala	Ile	Ala	Asp	Leu	Gly	Leu	Ala	Val	Arg	His	Asp	Ala	Val	Thr	Asp	
350					355					360					365	
ACC	ATA	GAC	ATT	GCT	CCA	AAT	CAG	AGG	GTG	GGG	ACC	AAA	CGA	TAC	ATG	1152
Thr	Ile	Asp	Ile	Ala	Pro	Asn	Gln	Arg	Val	Gly	Thr	Lys	Arg	Tyr	Met	
				370					375					380		
GCT	CCT	GAA	GTC	CTT	GAC	GAG	ACA	ATC	AAC	ATG	AAG	CAC	TTT	GAC	TCC	1200
Ala	Pro	Glu	Val	Leu	Asp	Glu	Thr	Ile	Asn	Met	Lys	His	Phe	Asp	Ser	
			385					390					395			
TTC	AAA	TGT	GCC	GAC	ATC	TAT	GCC	CTC	GGG	CTT	GTC	TAC	TGG	GAG	ATT	1248
Phe	Lys	Cys	Ala	Asp	Ile	Tyr	Ala	Leu	Gly	Leu	Val	Tyr	Trp	Glu	Ile	
		400					405					410				
GCA	CGA	AGA	TGC	AAT	TCT	GGA	GGA	GTC	CAT	GAA	GAC	TAT	CAA	CTG	CCG	1296
Ala	Arg	Arg	Cys	Asn	Ser	Gly	Gly	Val	His	Glu	Asp	Tyr	Gln	Leu	Pro	
	415					420					425					
TAT	TAC	GAC	TTA	GTG	CCC	TCC	GAC	CCT	TCC	ATT	GAG	GAG	ATG	CGA	AAG	1344
Tyr	Tyr	Asp	Leu	Val	Pro	Ser	Asp	Pro	Ser	Ile	Glu	Glu	Met	Arg	Lys	
430					435					440					445	
GTT	GTA	TGT	GAC	CAG	AAG	CTA	CGG	CCC	AAT	GTC	CCC	AAC	TGG	TGG	CAG	1392
Val	Val	Cys	Asp	Gln	Lys	Leu	Arg	Pro	Asn	Val	Pro	Asn	Trp	Trp	Gln	
				450					455					460		
AGT	TAT	GAG	GCC	TTG	CGA	GTG	ATG	GGA	AAG	ATG	ATG	CGG	GAG	TGC	TGG	1440
Ser	Tyr	Glu	Ala	Leu	Arg	Val	Met	Gly	Lys	Met	Met	Arg	Glu	Cys	Trp	
			465					470					475			
TAC	GCC	AAT	GGT	GCT	GCC	CGT	CTG	ACA	GCT	CTG	CGC	ATC	AAG	AAG	ACT	1488
Tyr	Ala	Asn	Gly	Ala	Ala	Arg	Leu	Thr	Ala	Leu	Arg	Ile	Lys	Lys	Thr	
		480					485					490				

bioRxiv preprint doi: <https://doi.org/10.1101/201606>; this version posted June 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CTG TCC CAG CTA AGC GTG CAG GAA GAT GTG AAG ATT TAAGCTGTTC 1534
 Leu Ser Gln Leu Ser Val Gln Glu Asp Val Lys Ile
 495 500 505

CTCTGCCTAC ACAAAGAACC TGGGCAGTGA GGATGACTGC AGCCACCGTG CAAGCGTCGT 1594

GGAGGCCTAT CCTCTTGTTT CTGCCCCGGCC CTCTGGCAGA GCCCTGGCCT GCAAGAGGGA 1654

CAGAGCCTGG GAGACGCGCG CACTCCCGTT GGGTTTGAGA CAGACACTTT TTATATTTAC 1714

CTCCTGATGG CATGGAGACC TGAGCAAATC ATGTAGTCAC TCAATGCCAC AACTCAAAC 1774

GCTTCAGTGG GAAGTACAGA GACCCAGTGC ATTGCGTGTG CAGGAGCGTG AGGTGCTGGG 1834

CTCGCCAGGA GCGGCCCCCA TACCTTGTGG TCCACTGGGC TGCAGGTTTT CCTCCAGGGA 1894

CCAGTCAACT GGCATCAAGA TATTGAGAGG AACCGGAAGT TTCTCCCTCC TTCCCGTAGC 1954

AGTCCTGAGC CACACCATCC TTCTCATGGA CATCCGGAGG ACTGCCCCCTA GAGACACAAC 2014

CTGCTGCCTG TCTGTCCAGC CAAGTGCGCA TGTGCCGAGG TGTGTCCAC ATTGTGCCTG 2074

GTCTGTGCCA CGCCCGTGTG TGTGTGTGTG TGTGTGAGTG AGTGTGTGTG TGTACACTTA 2134

ACCTGCTTGA GCTTCTGTGC ATGTGT 2160

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 505 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Ala Glu Ser Ala Gly Ala Ser Ser Phe Phe Pro Leu Val Val Leu
 1 5 10 15

Leu Leu Ala Gly Ser Gly Gly Ser Gly Pro Arg Gly Ile Gln Ala Leu
 20 25 30

Leu Cys Ala Cys Thr Ser Cys Leu Gln Thr Asn Tyr Thr Cys Glu Thr
 35 40 45

Asp Gly Ala Cys Met Val Ser Ile Phe Asn Leu Asp Gly Val Glu His
 50 55 60

His Val Arg Thr Cys Ile Pro Lys Val Glu Leu Val Pro Ala Gly Lys
 65 70 75 80

SEQUENCE 16

Pro	Phe	Tyr	Cys	Leu 85	Ser	Ser	Glu	Asp	Leu 90	Arg	Asn	Thr	His	Cys 95	Cys
Tyr	Ile	Asp	Phe 100	Cys	Asn	Lys	Ile	Asp 105	Leu	Arg	Val	Pro	Ser 110	Gly	His
Leu	Lys	Glu 115	Pro	Ala	His	Pro	Ser 120	Met	Trp	Gly	Pro	Val 125	Glu	Leu	Val
Gly	Ile 130	Ile	Ala	Gly	Pro	Val 135	Phe	Leu	Leu	Phe	Leu 140	Ile	Ile	Ile	Ile
Val 145	Phe	Leu	Val	Ile	Asn 150	Tyr	His	Gln	Arg	Val 155	Tyr	His	Asn	Arg	Gln 160
Arg	Leu	Asp	Met	Glu 165	Asp	Pro	Ser	Cys	Glu 170	Met	Cys	Leu	Ser	Lys 175	Asp
Lys	Thr	Leu	Gln 180	Asp	Leu	Val	Tyr	Asp 185	Leu	Ser	Thr	Ser	Gly 190	Ser	Gly
Ser	Gly	Leu 195	Pro	Leu	Phe	Val	Gln 200	Arg	Thr	Val	Ala	Arg 205	Thr	Ile	Val
Leu	Gln 210	Glu	Ile	Ile	Gly	Lys 215	Gly	Arg	Phe	Gly	Glu 220	Val	Trp	Arg	Gly
Arg 225	Trp	Arg	Gly	Gly	Asp 230	Val	Ala	Val	Lys	Ile 235	Phe	Ser	Ser	Arg	Glu 240
Glu	Arg	Ser	Trp	Phe 245	Arg	Glu	Ala	Glu	Ile 250	Tyr	Gln	Thr	Val	Met 255	Leu
Arg	His	Glu	Asn 260	Ile	Leu	Gly	Phe	Ile 265	Ala	Ala	Asp	Asn 270	Lys	Asp	Asn
Gly	Thr 275	Trp	Thr	Gln	Leu	Trp	Leu 280	Val	Ser	Asp	Tyr	His 285	Glu	His	Gly
Ser 290	Leu	Phe	Asp	Tyr	Leu	Asn 295	Arg	Tyr	Thr	Val	Thr 300	Ile	Glu	Gly	Met
Ile 305	Lys	Leu	Ala	Leu	Ser 310	Ala	Ala	Ser	Gly	Leu 315	Ala	His	Leu	His	Met 320
Glu	Ile	Val	Gly	Thr 325	Gln	Gly	Lys	Pro	Gly 330	Ile	Ala	His	Arg	Asp 335	Leu
Lys	Ser	Lys	Asn 340	Ile	Leu	Val	Lys	Lys 345	Asn	Gly	Met	Cys	Ala	Ile	Ala 350

Asp Leu Gly Leu Ala Val Arg His Asp Ala Val Thr Asp Thr Ile Asp
 355 360 365
 Ile Ala Pro Asn Gln Arg Val Gly Thr Lys Arg Tyr Met Ala Pro Glu
 370 375 380
 Val Leu Asp Glu Thr Ile Asn Met Lys His Phe Asp Ser Phe Lys Cys
 385 390 395 400
 Ala Asp Ile Tyr Ala Leu Gly Leu Val Tyr Trp Glu Ile Ala Arg Arg
 405 410 415
 Cys Asn Ser Gly Gly Val His Glu Asp Tyr Gln Leu Pro Tyr Tyr Asp
 420 425 430
 Leu Val Pro Ser Asp Pro Ser Ile Glu Glu Met Arg Lys Val Val Cys
 435 440 445
 Asp Gln Lys Leu Arg Pro Asn Val Pro Asn Trp Trp Gln Ser Tyr Glu
 450 455 460
 Ala Leu Arg Val Met Gly Lys Met Met Arg Glu Cys Trp Tyr Ala Asn
 465 470 475 480
 Gly Ala Ala Arg Leu Thr Ala Leu Arg Ile Lys Lys Thr Leu Ser Gln
 485 490 495
 Leu Ser Val Gln Glu Asp Val Lys Ile
 500 505

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1952 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mouse

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 187..1692

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AAGCGGCGGC AGAAGTTGCC GCGTGGTGC TCGTAGTGAG GGCGCGGAGG ACCCGGGACC

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

TGGGAAGCGG	CGGCGGGTTA	ACTTCGGCTG	AATCACAACC	ATTTGGCGCT	GAGCTATGAC		120									
AAGAGAGCAA	ACAAAAAGTT	AAAGGAGCAA	CCCGGCCATA	AGTGAAGAGA	GAAGTTTATT		180									
GATAAC	ATG	CTC	TTA	CGA	AGC	TCT	GGA	AAA	TTA	AAT	GTG	GGC	ACC	AAG	228	
	Met	Leu	Leu	Arg	Ser	Ser	Gly	Lys	Leu	Asn	Val	Gly	Thr	Lys		
	1				5					10						
AAG	GAG	GAT	GGA	GAG	AGT	ACA	GCC	CCC	ACC	CCT	CGG	CCC	AAG	ATC	CTA	276
Lys	Glu	Asp	Gly	Glu	Ser	Thr	Ala	Pro	Thr	Pro	Arg	Pro	Lys	Ile	Leu	
15					20					25					30	
CGT	TGT	AAA	TGC	CAC	CAC	CAC	TGT	CCG	GAA	GAC	TCA	GTC	AAC	AAT	ATC	324
Arg	Cys	Lys	Cys	His	His	His	Cys	Pro	Glu	Asp	Ser	Val	Asn	Asn	Ile	
				35					40					45		
TGC	AGC	ACA	GAT	GGG	TAC	TGC	TTC	ACG	ATG	ATA	GAA	GAA	GAT	GAC	TCT	372
Cys	Ser	Thr	Asp	Gly	Tyr	Cys	Phe	Thr	Met	Ile	Glu	Glu	Asp	Asp	Ser	
			50					55					60			
GGA	ATG	CCT	GTT	GTC	ACC	TCT	GGA	TGT	CTA	GGA	CTA	GAA	GGG	TCA	GAT	420
Gly	Met	Pro	Val	Val	Thr	Ser	Gly	Cys	Leu	Gly	Leu	Glu	Gly	Ser	Asp	
		65					70					75				
TTT	CAA	TGT	CGT	GAC	ACT	CCC	ATT	CCT	CAT	CAA	AGA	AGA	TCA	ATT	GAA	468
Phe	Gln	Cys	Arg	Asp	Thr	Pro	Ile	Pro	His	Gln	Arg	Arg	Ser	Ile	Glu	
	80					85					90					
TGC	TGC	ACA	GAA	AGG	AAT	GAG	TGT	AAT	AAA	GAC	CTC	CAC	CCC	ACT	CTG	516
Cys	Cys	Thr	Glu	Arg	Asn	Glu	Cys	Asn	Lys	Asp	Leu	His	Pro	Thr	Leu	
95					100					105					110	
CCT	CCT	CTC	AAG	GAC	AGA	GAT	TTT	GTT	GAT	GGG	CCC	ATA	CAC	CAC	AAG	564
Pro	Pro	Leu	Lys	Asp	Arg	Asp	Phe	Val	Asp	Gly	Pro	Ile	His	His	Lys	
				115					120					125		
GCC	TTG	CTT	ATC	TCT	GTG	ACT	GTC	TGT	AGT	TTA	CTC	TTG	GTC	CTC	ATT	612
Ala	Leu	Leu	Ile	Ser	Val	Thr	Val	Cys	Ser	Leu	Leu	Leu	Val	Leu	Ile	
			130					135					140			
ATT	TTA	TTC	TGT	TAC	TTC	AGG	TAT	AAA	AGA	CAA	GAA	GCC	CGA	CCT	CGG	660
Ile	Leu	Phe	Cys	Tyr	Phe	Arg	Tyr	Lys	Arg	Gln	Glu	Ala	Arg	Pro	Arg	
		145					150					155				
TAC	AGC	ATT	GGG	CTG	GAG	CAG	GAC	GAG	ACA	TAC	ATT	CCT	CCT	GGA	GAG	708
Tyr	Ser	Ile	Gly	Leu	Glu	Gln	Asp	Glu	Thr	Tyr	Ile	Pro	Pro	Gly	Glu	
	160					165					170					
TCC	CTG	AGA	GAC	TTG	ATC	GAG	CAG	TCT	CAG	AGC	TCG	GGA	AGT	GGA	TCA	756
Ser	Leu	Arg	Asp	Leu	Ile	Glu	Gln	Ser	Gln	Ser	Ser	Gly	Ser	Gly	Ser	
175					180					185					190	

GGC	CTC	CCT	CTG	CTG	GTC	CAA	AGG	ACA	ATA	GCT	AAG	CAA	ATT	CAG	ATG	804
Gly	Leu	Pro	Leu	Leu	Val	Gln	Arg	Thr	Ile	Ala	Lys	Gln	Ile	Gln	Met	
			195						200					205		
GTG	AAG	CAG	ATT	GGA	AAA	GGC	CGC	TAT	GGC	GAG	GTG	TGG	ATG	GGA	AAG	852
Val	Lys	Gln	Ile	Gly	Lys	Gly	Arg	Tyr	Gly	Glu	Val	Trp	Met	Gly	Lys	
			210					215					220			
TGG	CGT	GGA	GAA	AAG	GTG	GCT	GTG	AAA	GTG	TTC	TTC	ACC	ACG	GAG	GAA	900
Trp	Arg	Gly	Glu	Lys	Val	Ala	Val	Lys	Val	Phe	Phe	Thr	Thr	Glu	Glu	
		225					230					235				
GCC	AGC	TGG	TTC	CGA	GAG	ACT	GAG	ATA	TAT	CAG	ACG	GTC	CTG	ATG	CGG	948
Ala	Ser	Trp	Phe	Arg	Glu	Thr	Glu	Ile	Tyr	Gln	Thr	Val	Leu	Met	Arg	
		240				245					250					
CAT	GAG	AAT	ATT	CTG	GGG	TTC	ATT	GCT	GCA	GAT	ATC	AAA	GGG	ACT	GGG	996
His	Glu	Asn	Ile	Leu	Gly	Phe	Ile	Ala	Ala	Asp	Ile	Lys	Gly	Thr	Gly	
255					260					265					270	
TCC	TGG	ACT	CAG	TTG	TAC	CTC	ATC	ACA	GAC	TAT	CAT	GAA	AAC	GGC	TCC	1044
Ser	Trp	Thr	Gln	Leu	Tyr	Leu	Ile	Thr	Asp	Tyr	His	Glu	Asn	Gly	Ser	
				275					280					285		
CTT	TAT	GAC	TAT	CTG	AAA	TCC	ACC	ACC	TTA	GAC	GCA	AAG	TCC	ATG	CTG	1092
Leu	Tyr	Asp	Tyr	Leu	Lys	Ser	Thr	Thr	Leu	Asp	Ala	Lys	Ser	Met	Leu	
			290					295					300			
AAG	CTA	GCC	TAC	TCC	TCT	GTC	AGC	GGC	CTA	TGC	CAT	TTA	CAC	ACG	GAA	1140
Lys	Leu	Ala	Tyr	Ser	Ser	Val	Ser	Gly	Leu	Cys	His	Leu	His	Thr	Glu	
		305					310					315				
ATC	TTT	AGC	ACT	CAA	GGC	AAG	CCA	GCA	ATC	GCC	CAT	CGA	GAC	TTG	AAA	1188
Ile	Phe	Ser	Thr	Gln	Gly	Lys	Pro	Ala	Ile	Ala	His	Arg	Asp	Leu	Lys	
	320					325					330					
AGT	AAA	AAC	ATC	CTG	GTG	AAG	AAA	AAT	GGA	ACT	TGC	TGC	ATA	GCA	GAC	1236
Ser	Lys	Asn	Ile	Leu	Val	Lys	Lys	Asn	Gly	Thr	Cys	Cys	Ile	Ala	Asp	
335					340					345					350	
CTG	GGC	TTG	GCT	GTC	AAG	TTC	ATT	AGT	GAC	ACA	AAT	GAG	GTT	GAC	ATC	1284
Leu	Gly	Leu	Ala	Val	Lys	Phe	Ile	Ser	Asp	Thr	Asn	Glu	Val	Asp	Ile	
				355					360					365		
CCA	CCC	AAC	ACC	CGG	GTT	GGC	ACC	AAG	CGC	TAT	ATG	CCT	CCA	GAA	GTG	1332
Pro	Pro	Asn	Thr	Arg	Val	Gly	Thr	Lys	Arg	Tyr	Met	Pro	Pro	Glu	Val	
			370					375					380			
CTG	GAC	GAG	AGC	TTG	AAT	AGA	AAC	CAT	TTC	CAG	TCC	TAC	ATT	ATG	GCT	1380
Leu	Asp	Glu	Ser	Leu	Asn	Arg	Asn	His	Phe	Gln	Ser	Tyr	Ile	Met	Ala	
		385					390					395				

GAC ATG TAC AGC TTT GGA CTC ATC CTC TGG GAG ATT GCA AGG AGA TGT 1428
 Asp Met Tyr Ser Phe Gly Leu Ile Leu Trp Glu Ile Ala Arg Arg Cys
 400 405 410
 GTT TCT GGA GGT ATA GTG GAA GAA TAC CAG CTT CCC TAT CAC GAC CTG 1476
 Val Ser Gly Gly Ile Val Glu Glu Tyr Gln Leu Pro Tyr His Asp Leu
 415 420 425 430
 GTG CCC AGT GAC CCT TCT TAT GAG GAC ATG AGA GAA ATT GTG TGC ATG 1524
 Val Pro Ser Asp Pro Ser Tyr Glu Asp Met Arg Glu Ile Val Cys Met
 435 440 445
 AAG AAG TTA CGG CCT TCA TTC CCC AAT CGA TGG AGC AGT GAT GAG TGT 1572
 Lys Lys Leu Arg Pro Ser Phe Pro Asn Arg Trp Ser Ser Asp Glu Cys
 450 455 460
 CTC AGG CAG ATG GGG AAG CTT ATG ACA GAG TGC TGG GCG CAG AAT CCT 1620
 Leu Arg Gln Met Gly Lys Leu Met Thr Glu Cys Trp Ala Gln Asn Pro
 465 470 475
 GCC TCC AGG CTG ACG GCC CTG AGA GTT AAG AAA ACC CTT GCC AAA ATG 1668
 Ala Ser Arg Leu Thr Ala Leu Arg Val Lys Lys Thr Leu Ala Lys Met
 480 485 490
 TCA GAG TCC CAG GAC ATT AAA CTC TGACGTCAGA TACTTGTGGA CAGAGCAAGA 1722
 Ser Glu Ser Gln Asp Ile Lys Leu
 495 500
 ATTTACACAGA AGCATCGTTA GCCCAAGCCT TGAACGTTAG CCTACTGCCC AGTGAGTTCA 1782
 GACTTTCCTG GAAGAGAGCA CGGTGGGCAG ACACAGAGGA ACCCAGAAAC ACGGATTCAT 1842
 CATGGCTTTC TGAGGAGGAG AAAGTGTGTTG GGTAAGTTGT TCAAGATATG ATGCATGTTG 1902
 CTTTCTAAGA AAGCCCTGTA TTTTGAATTA CCATTTTTTTT ATAAAAAAAAA 1952

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 502 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Leu Leu Arg Ser Ser Gly Lys Leu Asn Val Gly Thr Lys Lys Glu
 1 5 10 15
 Asp Gly Glu Ser Thr Ala Pro Thr Pro Arg Pro Lys Ile Leu Arg Cys
 20 25 30

Lys Cys His His His Cys Pro Glu Asp Ser Val Asn Asn Ile Cys Ser
 35 40 45
 Thr Asp Gly Tyr Cys Phe Thr Met Ile Glu Glu Asp Asp Ser Gly Met
 50 55 60
 Pro Val Val Thr Ser Gly Cys Leu Gly Leu Glu Gly Ser Asp Phe Gln
 65 70 75 80
 Cys Arg Asp Thr Pro Ile Pro His Gln Arg Arg Ser Ile Glu Cys Cys
 85 90 95
 Thr Glu Arg Asn Glu Cys Asn Lys Asp Leu His Pro Thr Leu Pro Pro
 100 105 110
 Leu Lys Asp Arg Asp Phe Val Asp Gly Pro Ile His His Lys Ala Leu
 115 120 125
 Leu Ile Ser Val Thr Val Cys Ser Leu Leu Leu Val Leu Ile Ile Leu
 130 135 140
 Phe Cys Tyr Phe Arg Tyr Lys Arg Gln Glu Ala Arg Pro Arg Tyr Ser
 145 150 155 160
 Ile Gly Leu Glu Gln Asp Glu Thr Tyr Ile Pro Pro Gly Glu Ser Leu
 165 170 175
 Arg Asp Leu Ile Glu Gln Ser Gln Ser Ser Gly Ser Gly Ser Gly Leu
 180 185 190
 Pro Leu Leu Val Gln Arg Thr Ile Ala Lys Gln Ile Gln Met Val Lys
 195 200 205
 Gln Ile Gly Lys Gly Arg Tyr Gly Glu Val Trp Met Gly Lys Trp Arg
 210 215 220
 Gly Glu Lys Val Ala Val Lys Val Phe Phe Thr Thr Glu Glu Ala Ser
 225 230 235 240
 Trp Phe Arg Glu Thr Glu Ile Tyr Gln Thr Val Leu Met Arg His Glu
 245 250 255
 Asn Ile Leu Gly Phe Ile Ala Ala Asp Ile Lys Gly Thr Gly Ser Trp
 260 265 270
 Thr Gln Leu Tyr Leu Ile Thr Asp Tyr His Glu Asn Gly Ser Leu Tyr
 275 280 285
 Asp Tyr Leu Lys Ser Thr Thr Leu Asp Ala Lys Ser Met Leu Lys Leu
 290 295 300

866760
 447600

Ala Tyr Ser Ser Val Ser Gly Leu Cys His Leu His Thr Glu Ile Phe
 305 310 315 320
 Ser Thr Gln Gly Lys Pro Ala Ile Ala His Arg Asp Leu Lys Ser Lys
 325 330 335
 Asn Ile Leu Val Lys Lys Asn Gly Thr Cys Cys Ile Ala Asp Leu Gly
 340 345 350
 Leu Ala Val Lys Phe Ile Ser Asp Thr Asn Glu Val Asp Ile Pro Pro
 355 360 365
 Asn Thr Arg Val Gly Thr Lys Arg Tyr Met Pro Pro Glu Val Leu Asp
 370 375 380
 Glu Ser Leu Asn Arg Asn His Phe Gln Ser Tyr Ile Met Ala Asp Met
 385 390 395 400
 Tyr Ser Phe Gly Leu Ile Leu Trp Glu Ile Ala Arg Arg Cys Val Ser
 405 410 415
 Gly Gly Ile Val Glu Glu Tyr Gln Leu Pro Tyr His Asp Leu Val Pro
 420 425 430
 Ser Asp Pro Ser Tyr Glu Asp Met Arg Glu Ile Val Cys Met Lys Lys
 435 440 445
 Leu Arg Pro Ser Phe Pro Asn Arg Trp Ser Ser Asp Glu Cys Leu Arg
 450 455 460
 Gln Met Gly Lys Leu Met Thr Glu Cys Trp Ala Gln Asn Pro Ala Ser
 465 470 475 480
 Arg Leu Thr Ala Leu Arg Val Lys Lys Thr Leu Ala Lys Met Ser Glu
 485 490 495
 Ser Gln Asp Ile Lys Leu
 500

2025-01-24 14:00:00

- (2) INFORMATION FOR SEQ ID NO: 19:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GCGGATCCTG TTGTGAAGGN AATATGTG

28

- (2) INFORMATION FOR SEQ ID NO: 20:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

GCGATCCGTC GCAGTCAAAA TTTT

24

- (2) INFORMATION FOR SEQ ID NO: 21:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

GCGGATCCGC GATATATTAA AAGCAA

26

2005-07-27 14:00:00

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: cDNA

- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: YES

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

20

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: cDNA

- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

37

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: cDNA

- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

26

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AACACCGGGC CGGCGATGAT

20

(2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 6 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Gly Xaa Gly Xaa Xaa Gly
 1 5

(2) INFORMATION FOR SEQ ID NO: 27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 6 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

Asp Phe Lys Ser Arg Asn
 1 5

856T60" 44T6E060

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Asp Leu Lys Ser Lys Asn

1

5

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

Gly Thr Lys Arg Tyr Met

1

5

SUBMITTED